No more hot air

European heads of state gathered in Brussels today must show they are serious about climate change.

When the European Union's 27 heads of state meet today, they are expected to endorse an energy policy for the continent whose implementation will provide the ultimate test of Europe's ability to meet the challenge of climate change.

European leaders such as Tony Blair and Angela Merkel have talked a good game on the climate-change issue for many years. But Europe's actions to confront stubborn, continued growth in greenhouse-gas emissions have been far less impressive. Today's rhetoric will doubtless imply that this is about to change. However, it is the actual implementation of the plan, rather than the flourish with which it is presented, that will show whether the European Union (EU) is serious about cutting emissions. Sceptics on both sides of the Atlantic think it isn't — and will be watching closely.

Most Europeans want their elected leaders to rise to the challenge and take effective action. They understand that this will involve difficult lifestyle choices regarding, for example, the size of the cars they drive and the amount of energy they waste at home. People are prepared to make these choices provided there is a sense that the burden is being fairly shared.

The heads of state will today discuss an action plan that was drawn up by the European Commission in January and endorsed at a meeting of EU environment ministers last month. It commits the EU as a whole to mandatory reductions in greenhouse-gas emissions of 20–30% from 1990 levels by 2020 (see *Nature* **445,** 234–235; 2007).

The plan would stipulate a 20% reduction in emissions and allow for a more arduous 30% cut if other developed countries (primarily the United States) agree to take mandatory action to reduce their own emissions. It also incorporates a number of secondary, legally binding targets, such as the generation of 20% of electricity from renewable sources by 2020.

Given doubts over the true potential of renewable energy sources and other uncertainties, the mandatory nature of the proposal has faced strenuous opposition. However, voluntary measures have done little to change behaviour over the past two decades, so the EU's leaders must now endorse a package of mandatory measures — if only to back up their own well-worn rhetoric on the importance of the climate-change issue. Such mandatory commitments are clearly needed to provide the incentives that will increase research and investment in clean-energy technologies, energy efficiency and the other steps needed to reverse the growth in emissions.

The national circumstances of the leaders gathered in Brussels are varied. Italy, Spain and Portugal, for example, are failing to meet even the modest targets allocated to them as their share of the EU's commitment under the Kyoto Protocol, as the new-found fashion for air-conditioning sends their electricity usage through the roof. These countries, along with the poorer member states in the east, are likely to push for the largest and most intensive energy users — Germany, Britain and France — to bear most of the burden imposed by the energy plan. However, the leaders at the meeting realize how wretched they will all look if no agreement is reached, and can therefore be expected to find one.

"Most Europeans want their elected leaders to rise to the challenge and take effective action on climate change."

The current logjam in international climate diplomacy will only be broken by decisive European leadership on the issue. Anything less will recall the grotesque embarrassment that accompanied EU hesitation in the Balkans in the 1990s, when only US intervention prevented a bloodbath.

The climate crisis is a different type of challenge, but is no less momentous for all that. It provides a timely opportunity for the EU to prove that it is capable of leading the world in an issue of global importance. This will require technological innovation, societal backing, economic adaptability and, above all, political will. There is no reason — yet — to conclude that the EU isn't equal to the task. ■

Tackling tuberculosis

There is a dangerous gulf between the global programmes to fight AIDS and TB.

Tuberculosis (TB) is one of the world's most lethal diseases, and is expected to kill almost 2 million people this year. But it has a relatively low public profile. Worse still, efforts to research, prevent and treat the disease are entangled in an unseemly rivalry with corresponding approaches to AIDS.

TB has been with us for centuries, and has probably already claimed many of the people in the West who are genetically susceptible to it. By far the greatest prevalence today is in Africa and India, where susceptibility to TB often goes hand-in-hand with HIV infection (see *Nature Med.* **13,** 263; 2007 and www.nature.com/news/specials/tb).

Yet the response to the march of the disease in poor countries has been patchy. The World Health Organization (WHO) has led a moderately successful effort to offer basic treatments to as many people as possible, using 40-year-old drugs of limited effectiveness. The US National Institute of Allergy and Infectious Diseases sponsors by far the world's largest TB research programme, and the Bill & Melinda Gates Foundation is making a determined push to reboot efforts to translate research into clinically useful drugs and vaccines.

However, all this activity is modest in relation to the scale of the problem. TB research programmes are estimated to be worth about $300 million worldwide — less than one-tenth of the amount devoted to AIDS. Important questions, including how to develop better ways

of detecting the disease cheaply and reliably, and how to assess the effectiveness of new drug combinations, go unanswered.

But it isn't just the lack of funding that makes TB researchers feel like the poor relations in the global health family. Difficulties in working alongside the better-funded efforts to tackle AIDS permeate many aspects of TB prevention, treatment and research at all levels, from the clinics of South Africa's most impoverished townships to the headquarters of the WHO in Geneva.

Treatment for TB is commonly dispensed in developing countries through long-established public-health clinics, whereas AIDS treatment often comes through separate, dedicated facilities. This division does nothing for the patients suffering from both diseases.

As for research, there has been inadequate collaboration between the relatively small number of scientists who devote their careers to studying TB and the far larger community working on AIDS. Yet both groups could surely benefit from each other, especially with regard to improving their understanding of the interaction between the two diseases.

But it is perhaps at the bureaucratic level that this division is most pronounced. The offices dealing with AIDS and TB at the WHO, for example, have historically enjoyed a difficult relationship characterized by rivalry rather than cooperation.

Steps are being taken to improve this sorry state of affairs. Treatment is successfully being integrated on the ground in South Africa and elsewhere, usually one clinic at a time. Researchers from the two fields are working together to study 'immune reconstitution syndrome', a little-understood phenomenon that affects those taking drugs for both TB and AIDS. And at the WHO, outmoded blueprints for the diagnosis and treatment of TB are being revised, with the assistance of staff with backgrounds in HIV treatment. Thanks to the intervention of the Gates Foundation and others, drug and vaccine candidates for TB are entering trials at a reasonably healthy rate for the first time in decades.

There is an overriding need for greater collaboration between AIDS and TB prevention, treatment and research, and this should be implemented at the grass roots wherever possible. But this common-sense remedy isn't, on its own, going to overturn a deeply ingrained division that has taken shape over many years.

Global health provision only really changes when leadership is forthcoming, not just from international organizations but from governments, specifically those that are in a position to lead — in this case, those of the United States, the European Union, India and South Africa. World Tuberculosis Day on 24 March seeks to draw attention to the disease, and provides an opportunity for governments to do just that, by acknowledging the problem and stating what they intend to do about it. ■

The ends of the Earth

International Polar Year 2007 can leave an imprint.

The last time there was an International Polar Year (IPY), the world was a very different place. In 1957, at the height of the cold war, the poles were less a place for intriguing scientific discoveries than for political manoeuvring between the Soviet Union and the United States. Yet 1957–58 marked the third International Polar Year (the first two were in 1882 and 1932) and the first International Geophysical Year — a significant landmark, with hindsight, for global scientific collaboration.

The International Geophysical Year yielded several great scientific discoveries, many of which were inexorably linked to cold-war imperatives. James Van Allen discovered the belts of radiation surrounding Earth with the first US satellite, Explorer I, sent up in response to the Soviet Union's Sputnik. And the US nuclear submarine *Nautilus*, on a top-secret voyage, became the first vessel to visit the North Pole under the ice.

This time round, the political context of the IPY is dominated by climate change. Tasks for the coming year include taking detailed measurements of melting sea ice in the Arctic (see page 133) and a hunt for the best records yet of past climate change in the Antarctic (see page 126).

But scientists seeking support for these missions have got off to a rocky start, at least in the United States, as the result of a budget impasse that briefly froze spending for the National Science Foundation at 2006 levels. By January, last-minute negotiations had yielded an extra $334 million for the agency, including money it needs to grant IPY proposals in the coming year. But many of the scientists involved are still awaiting confirmation that their projects will go ahead, and their frustration is becoming palpable.

Elsewhere, the outlook is more certain. Canada, for instance, has pulled together all the funding for a large international programme to study the circumpolar flaw lead, an area of water that separates the bulk of the Arctic sea ice from the ice at the coast. Canada has also taken a welcome lead in involving indigenous peoples in its research programmes. The interdisciplinary ArcticNet project, based at the Université Laval in Quebec, aims to disseminate information gleaned about changes in the polar regions to the communities that are most directly affected.

These international efforts may alleviate some of the problems that have long plagued Arctic research, such as the decline of meteorological monitoring stations. The collapse of the Soviet Union has caused numerous observing stations to close. One target for the IPY is to upgrade some key sites and monitor them over the long term; the US National Oceanic and Atmospheric Administration, for instance, is upgrading its climate-monitoring laboratory at Barrow, Alaska, and plans to do the same for stations in Eureka, Canada, and Tiksi, Russia. Such observatories are the only way to move forward with collecting the long-term data needed to monitor climate change.

> **"The polar year represents the best chance to get climate-monitoring networks up and running."**

The polar-year celebration represents, among other things, the best chance to get these climate-monitoring networks up and running. Let's hope that the funding difficulties are ironed out in time, and that this opportunity is used to produce a legacy worthy of International Polar Year 2007. ■

RESEARCH HIGHLIGHTS

Holding up the ice

Science doi:10.1126/science.1138393 and doi:10.1126/science.1138396 (2007)

Antarctic ice may be protected from rising sea levels by rocky wedges of debris that act as 'sandbags', a survey of one of the continent's major ice flows suggests.

A team led by Sridhar Anandakrishnan of Pennsylvania State University surveyed the final stretches of the 500-kilometre-long Whillans Ice Stream, including the 'grounding line' where the glacier passes from land to sea. Underneath the grounding line, debris deposited by the glacier has formed a wedge up to 31 metres thick. Models described in a second paper predict that the wedge will stabilize the ice stream — preventing sea water from seeping under the glacier and accelerating melting until sea levels have risen by several metres.

S. ANANDAKRISHNAN

BIOCHEMISTRY
A first for vitamins

Nature Chem. Biol. doi:10.1038/nchembio867 (2007)

The discovery of a vitamin attached to adenosine triphosphate (ATP), the molecule that fuels biochemical reactions, has taken researchers by surprise.

Lucien Bettendorff at the University of Liège in Belgium and his colleagues report that vitamin B1, or thiamine, occurs naturally in complex with ATP — the first vitamin known to do so. The researchers found that the bacterium *Escherichia coli* produces the complex when starved of glucose. The team also identified the complex in yeast, plant and mammalian cells, suggesting that it has an as yet unknown basic function in cell metabolism, perhaps in signalling.

NEUROBIOLOGY
Neurons get connected

Neuron **53**, 639–647 (2007)

Researchers have devised a way of identifying all the brain cells that connect to one

ELSEVIER

particular brain cell. The technique should help neuroscientists to understand how networks of neurons are wired together.

Current 'tracers' can spread through chains of neural connections, but are not specific enough to show just one step of a network. Now Ian Wickersham at the Salk Institute for Biological Studies in La Jolla, California, and his colleagues have modified the rabies virus, a potent tracer, so that it can do just that.

They disabled genes that help the virus to jump between neurons, then delivered the modified virus to a neuron that can make the protein the virus needs to spread. This meant the virus could proceed to immediately connected neurons, but no further.

The researchers demonstrated the method in culture (green dots in the picture below are neurons that connect to the neurons shown in yellow); they say that it should also work *in vivo*.

GENETICS
Cycle of all life

Proc. Natl Acad. Sci. USA **104**, 2939–2944 (2007)

Geneticists have mapped the genes that underpin the cell cycle in one species of Archaea, in a bid to discover whether cell growth and division involve similar genes across all branches of the tree of life.

Magnus Lundgren and Rolf Bernander of Uppsala University, Sweden, studied the heat-loving microbe *Sulfolobus acidocaldarius*. By sampling the abundance of RNA from various genes at different points in the cell cycle, they built up a catalogue of more than 160 genes that are activated at specific points.

Comparing the genes that control the cell cyles in Archaea, bacteria and eukaryotes — which diverged from each other early on — could provide information about how the cycle evolved, and identify key processes within it.

AGEING
New ideas on ageing theory

Nature Genet. doi:10.1038/ng1988 (2007)

A controversial theory of ageing is rebutted by a study that finds no link between the accumulation of mutations in energy-producing organelles called mitochrondria and symptoms of old age.

Researchers have previously speculated, on the back of work that documented the accumulation of mitochondrial mutations in mice over time, that mitochondrial failure contributes to ageing. Lawrence Loeb of the University of Washington, Seattle, and his colleagues have now measured the rate of single-base-pair mitochondrial mutations in mice to be tenfold lower than reported in earlier studies. Loeb's team also found that a mouse mutant that accumulates mutations at 500 times the rate of a normal mouse does not show signs of rapid ageing.

DRUG DISCOVERY
Tumour target

Proc. Natl Acad. Sci. USA **104**, 3478–3483 (2007)

Researchers have developed an animal model to test cancer drugs that target a tumour's blood supply.

These drugs prevent blood-vessel growth in tumours by inhibiting a protein known as VEGF-A. Several drugs have been approved to treat specific forms of lung and colorectal cancer by this method. But although the
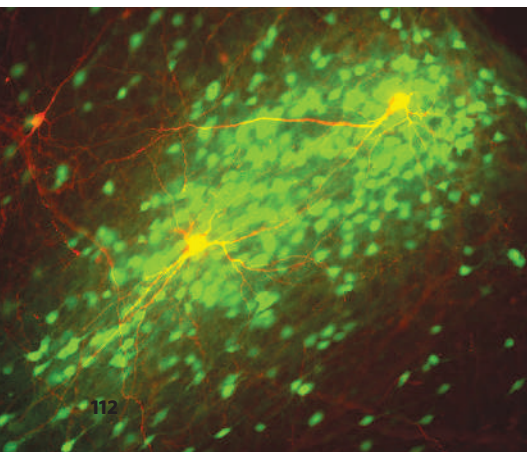
drugs blocked tumour blood-vessel growth in humans, they failed to do so in rodents, complicating the testing process.

Now Napoleone Ferrara and his colleagues at Genentech in South San Francisco, California, have engineered mice to produce a human form of VEGF-A. They have also tested new candidate drugs and were able to weed out some that bound tightly to their VEGF target *in vitro*, but did not reduce tumour growth in engineered mice.

BIOCHEMISTRY
Lighting rods

J. Am. Chem. Soc. doi:10.1021/ja063887t (2007)
A structure that mimics the exquisite light-harvesting machinery of nature has been made from a virus's self-assembling shell.

Some photosynthetic bacteria capture and concentrate light's energy in barrel-shaped protein structures, which have light-absorbing molecules known as chromophores stacked around their rim. Matthew Francis and his co-workers at the University of California, Berkeley, copied this design by attaching chromophores to proteins from the coat of the tobacco mosaic virus, which self-assembles into stacks of disks or rods.

Some of the chromophores absorb light; others act as 'acceptors', which gather and re-emit energy from many of the 'donors'. Energy concentrated in this way might drive photocatalysis or improve photovoltaic cells.

CANCER BIOLOGY
Disruptive influence

Science doi:10.1126/science.1137999 (2007)
In some cancers, the regulation of gene expression by short pieces of RNA known as microRNAs is disrupted. But it has been difficult to pin down the mechanisms
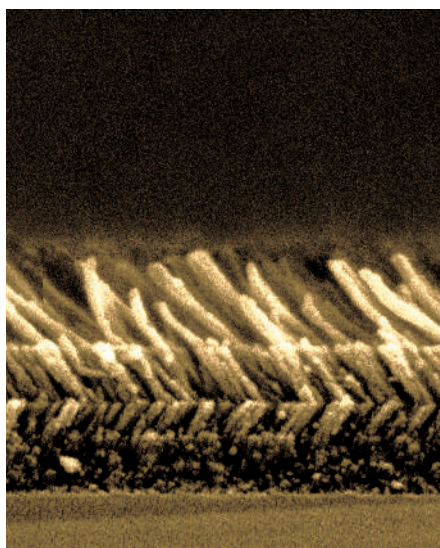
involved. David Bartel of the Whitehead Institute for Biomedical Research in Cambridge, Massachusetts, and his colleagues present a clear example: it involves one microRNA acting on one target.

The researchers examined a protein called High Mobility Group A2 (Hmga2). Truncated forms of this protein are known to trigger cancer. Bartel's team found that the messenger RNA encoding the truncated protein is missing a region that would usually bind a microRNA known as *let-7*. The team showed that this interaction represses accumulation of the protein, and that disrupting it spurs cell growth in culture and causes tumours in mice.

OPTICS
No reflection

Nature Photon. **1**, 176–179 (2007)
A near-perfect anti-reflective coating that can make even shiny materials appear dull and dark is described in the current issue of

Nature Photonics. The coating, which reflects less than 0.5% of the visible light falling on it, could have applications in devices such as solar cells.

Light gets reflected when it passes between materials in which it travels at different speeds — measured by the material's 'refractive index'. To minimize these reflections, Fred Schubert of Rensselaer Polytechnic Institute in Troy, New York state, and his colleagues designed a coating that has five layers (pictured left), with refractive indices that increase in steps. The two uppermost layers, arrays of silica nanorods, have refractive indices close to that of air. The lower layers, made from titanium dioxide, have indices that tend towards that of the material beneath.

PHYSICS
The test of time

Phys. Rev. Lett. doi:10.1103/PhysRevLett.98.093001 (2007)
Researchers have, for the first time, captured radium atoms in a magneto-optical trap, laying the groundwork for a fundamental test of time's properties.

Jeffrey Guest and a team at Argonne National Laboratory in Illinois trapped radium-225 and radium-226 atoms by cooling them with a combination of lasers and magnetic coils. The team was aided by the ubiquitous thermal radiation from their room-temperature apparatus, which helped to confine the atoms.

Radium atoms may eventually be used to test 'time-reversal symmetry', which says that there should be no way to determine the direction of time from observations of the atoms' properties. If this symmetry were violated, it might help to explain the imbalance of matter and antimatter in the Universe.

JOURNAL CLUB

Jeffery W. Kelly
The Scripps Research Institute, La Jolla, California, USA

A biochemist considers whether protein misfolding plays a part in type II diabetes.

Much of my research is on cellular protein folding, and in particular on how protein misfolding or protein aggregation causes disease. My group has developed therapies for a spectrum of misfolding diseases, most of which are associated with neurodegeneration, such as Alzheimer's.

But we are beginning to appreciate that therapies that affect protein folding could have a role in treating a much wider spectrum of diseases than is currently realized.

A compelling article from Gokhan Hotamisligil and his colleagues at Harvard University (U. Özcan *et al. Science* **313**, 1137–1140; 2006) presents one example. They found that mice that are both obese and diabetic benefit from treatment with drugs that enhance protein folding.

Their experiment was motivated by observations that linked obesity and diabetic insulin resistance to stress in the endoplasmic reticulum (ER), a compartment in cells where a third of all proteins are folded.

The researchers gave their fat, diabetic mice chemicals that enhance protein folding in the ER. The effect was notable: the mice's blood-sugar levels fell, they showed increased glucose tolerance and reduced lipid accumulation in the liver.

This suggests to me that protein misfolding may be at the heart of type II diabetes, the age-related disease for which these mice are a model.

Folding of the insulin receptor is inefficient. So it seems reasonable to speculate that cells could become insulin-resistant because of compromised insulin-receptor folding in the ER.

We may find, as we develop more selective small molecules to enhance ER folding, that we discover other disorders that can be treated in this way.

NEWS

Western states launch carbon scheme

If the federal government doesn't get its act together on carbon trading, individual US states will come together to decide policy instead. That was the clear message when the governors of five western states — Washington, Oregon, California, Arizona and New Mexico — signed an agreement on 26 February to create their own carbon cap-and-trade programme. The move parallels a similar scheme developed by northeastern states; together the two regions account for nearly a quarter of US carbon emissions.

As California's governor Arnold Schwarzenegger puts it: "This agreement shows the power of states to lead our nation in addressing climate change."

The western group is moving relatively quickly. It also intends to go further than the northeastern scheme, the ten-state Regional Greenhouse Gas Initiative (RGGI), which focuses on the electricity industry. "Within 18 months, we want to return with a design for a multi-sector, market-based system," says Dan Skopec, an undersecretary at the California Environmental Protection Agency. Each state will set its own caps.

California was already devising a market approach to managing emissions. But some worried that regulated companies would leave the state to avoid being capped. The new scheme, the Western Regional Climate Action Initiative (WRCAI), has nearly all the states that border California on board, so such leakage seems unlikely. The exception is Nevada, but Skopec says he still hopes it will join.

In the electricity sector, the WRCAI may

The United States is seeing a wave of regional climate initiatives.

operate differently to RGGI. Starting in 2009, RGGI will cap companies that generate energy, and they must then either ramp down operations, invest in cleaner technologies, or buy offsets to meet their targets. California had favoured the idea of capping the distributors who sell electricity to the customer, to cover all electricity sold in the region no matter where it was produced. But with neighbouring states on board as well, the scheme may stick to capping the generators.

On a national level, the Democratic Congress is holding hearings and proposing bills on emissions at a breakneck pace, but even with the new initiative, the crucial 60 votes needed in the Senate look unlikely to be found. As such schemes become more popular, however, pressure will mount on the federal government to come up with a cap-and-trade scheme or risk having states do it for them.

According to the Pew Center on Global Climate Change, if all the states in the two regional schemes counted as a country they would be the world's fourth largest emitter. "At some point you have critical mass. At some point you have enough of a system that it just becomes the standard, and states sign on voluntarily

Scriptural violence can foster aggression

There once was a man and his concubine from the Israeli tribe of Ephraim who were travelling in the land of Benjamin, another Israeli tribe. As the couple dined in the city of Gibeah, a mob assembled outside and pounded on the door. The mob captured the concubine, then raped and beat her to death. The man collected her corpse the next day and travelled home. The other tribes of Israel were outraged at the crime, assembled an army and razed several Benjamite cities, killing every man, woman, child and animal they could.

Around 500 students recently read a version of this story, which is based on a passage from the Old Testament, as part of a psychological study. For half of the participants the tale contained an additional passage: when the man returned home, his tribe prayed to God and asked what they should do. God commanded the tribe to "take arms against their brothers and chasten them before the Lord".

After reading the story, the students participated in another exercise intended to measure aggression. About half of the study participants came from Brigham Young University, a religious university in Provo, Utah, and almost all were members of the Church of Jesus Christ of Latter-Day Saints. The other half came from the Free University in Amsterdam. Only 50% of the Dutch group believed in God and 27% in the Bible. But for both groups — whether the students were based in the Netherlands or the United States, and believed in God or not — the trend was the same: those who were told that God had sanctioned the violence against the Israelite were more likely to act aggressively in the subsequent exercise.

The study is indicative of a growing interest among psychologists and sociologists in the origins of religious violence. That subject was taboo until recently for many psychologists, and past research tended to focus on the role of religion

"If violence is presented as the authoritative voice of God, it can increase the possibility of more violence."

in psychological healing. But heightening concern about religious terrorism has pushed negative uses of religion to the forefront. "People often use God as a justification for committing violent acts," says Brad Bushman, a social psychologist at the University of Michigan in Ann Arbor and lead author of the study. "And that just bothers me, I guess."

The results of Bushman's study, to be published in the March issue of *Psychological Science*, do not indicate that religious people are more aggressive than non-religious people (B. J. Bushman *et al. Psychol. Sci.* **18,** 204–207; 2007). Furthermore, the story us ed was an isolated example of scriptural violence taken out of context, and thus does not reflect the experience of reading the Bible as a whole. But it does suggest that selective exposure to violent passages in a scriptural canon can promote aggression.

That response probably reflects a long-standing finding in psychology that people respond more aggressively to a depiction of
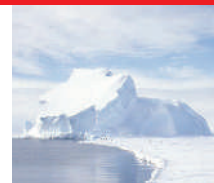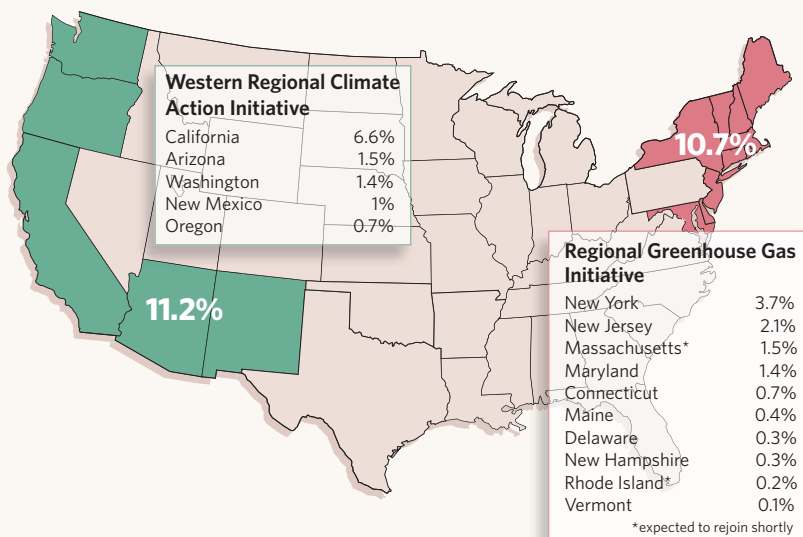
REGULATING GREENHOUSE GASES

A new initiative to control carbon dioxide emissions in the US west complements an earlier initiative in the east. Here's a look at the states that are participating, and the percentage of total US carbon emissions they make up.

Western Regional Climate Action Initiative

California	6.6%
Arizona	1.5%
Washington	1.4%
New Mexico	1%
Oregon	0.7%

11.2%

10.7%

Regional Greenhouse Gas Initiative

New York	3.7%
New Jersey	2.1%
Massachusetts*	1.5%
Maryland	1.4%
Connecticut	0.7%
Maine	0.4%
Delaware	0.3%
New Hampshire	0.3%
Rhode Island*	0.2%
Vermont	0.1%

*expected to rejoin shortly

or the federal government will just adopt it," says Véronique Bugnion, the Washington DC-based director of research for emissions trading consultants Point Carbon.

To help reach that critical mass, the two schemes may link up, and perhaps even join forces with the European Union (EU) Emission Trading Scheme. "The bigger the market, the more cost-effective reductions are available," says John Larsen, a research analyst at the

World Resources Institute, an environmental group in Washington DC. "It will be interesting to see how they move forward and who they talk to."

Skopec says that California is already moving towards this, but there are some potential hitches. One problem is the different way the schemes are regulated. For example, if more states joined the two US groups, electricity generated in a RGGI state and exported to a state in the WRCAI, which may regulate distributors, could potentially be counted twice.

And differences in how tough the caps are could set up a lopsided situation in which companies in low-cap areas keep buying credits from high-cap areas. "You have to link equivalents," says Bugnion. "If one system is more stringent than another, the credits will only flow one way." There's also the problem that the EU trading scheme operates under the aegis of the Kyoto Protocol, a treaty that the United States hasn't ratified and that individual US states cannot join.

A looser form of linking may be easier. The RGGI plan already allows the purchase of off-set credits from projects certified by the Clean Development Mechanism, an apparatus of the Kyoto Protocol. A shared pool of available offsets can bring prices in the separate markets closer together, even if they aren't linked directly. ■

Emma Marris

---

violence that they feel is justified, says Robert Ridge, a social psychologist at Brigham Young University and a co-author of the study.

Sociologist Mark Juergensmeyer of the University of California, Santa Barbara, says his research has also pointed to the motivational power of scriptural violence, but that the context of the message is key. "If violence is presented as the authoritative voice of God, it can increase the possibility of more violence," says Juergensmeyer. "But everything depends on how it is presented." The same passage placed in a non-threatening, historical context might not promote aggression, he argues.

Nevertheless, when scriptural violence is used to promote hostility, it is extremely effective, Juergensmeyer adds. Invoking religious justification allows a political leader to believe in promises of immortality and

spiritual rewards that can be powerful motivators. "Religion is not the problem," he says. "But it can make a secular problem worse."

People often choose to ignore the violent side to religion, says John Hall, a sociologist at the University of California, Davis, and they tend to dismiss those who commit religiously inspired violence as members of the fringe. "There are built-in cultural lenses that we use to dissociate religion from violence," he says. "When we see religious movements that are prophetically inspired and engaged in violence, there's a cultural tendency to say 'oh, they're not really religious'. "

That view represents a misleading, selective interpretation of most religious canons, agrees theologian Hector Avalos of Iowa State University in Ames. "People who choose the violent interpretation are no less arbitrary than those who choose the peaceful

God's sanction can be a motivator for aggression.

CORBIS

one," he says. Avalos has proposed a radical solution to theologically inspired violence — cut the violent passages out of the scripture.

It's a wildly controversial idea that ought not to be, he says, because spiritual leaders effectively

do that on a regular basis. "A lot of churches have a series of passages that they read during the year," says Avalos. "And usually they don't choose the passages involving genocide."

Heidi Ledford

OUR SOLAR SYSTEM LAST WEEK

Over the past few days, Earthlings have been treated to some stunning images from across the Solar System. Jupiter and Mars were snapped by spacecraft on their way to more distant destinations, the gap between Earth and the Sun was captured by the STEREO mission, and the Cassini spacecraft collated months of hard work to give us fresh views of Saturn and its moon Titan.
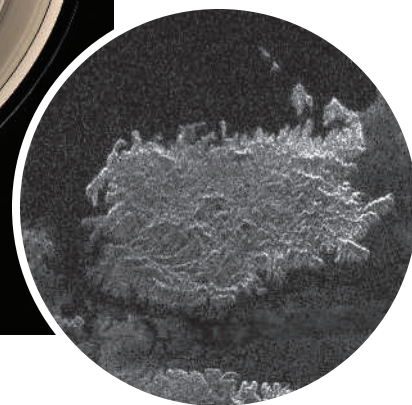
NASA/JPL/SPACE SCI. INST.

NASA/JPL
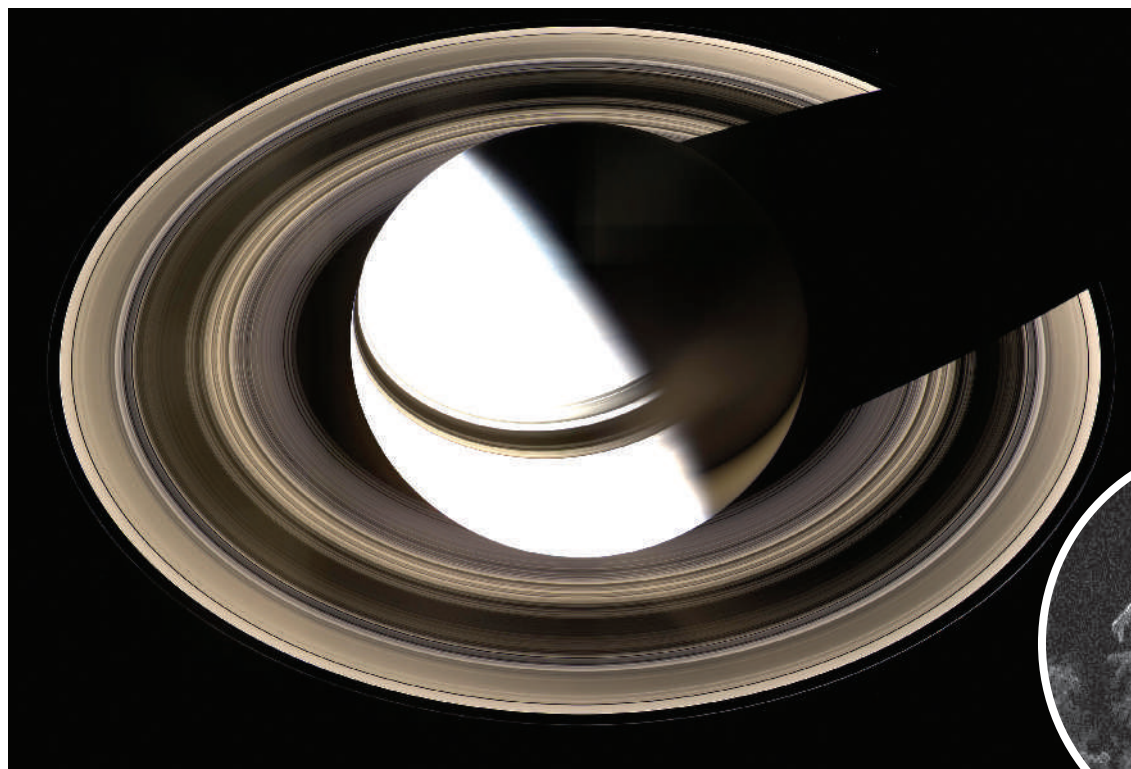
◀ SATURN'S SHADOW

NASA's Cassini spacecraft has spent months circling ever higher above Saturn. This picture of the planet and its rings (left), complete with shadow, is made from 36 images taken over 2.5 hours. Cassini's radar, meanwhile, has captured an image of an island in the middle of a lake on Saturn's moon Titan (inset).

NASA/JOHNS HOPKINS UNIV./SOUTHWEST RES. INST.
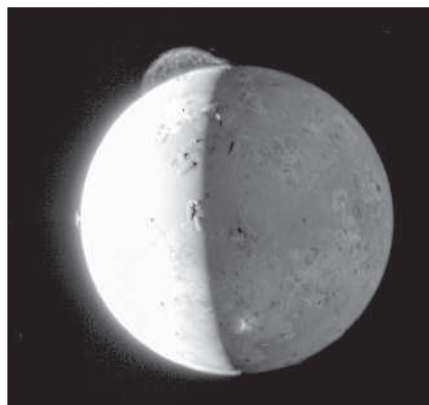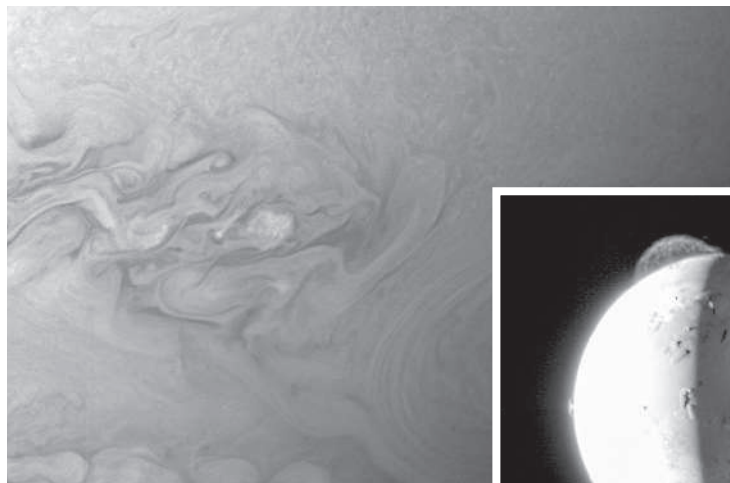
◀ STORM CLOUDS AND VOLCANIC PLUMES

The culmination of three storms on Jupiter, known as the little red spot, is seen swirling over the surface of the planet (far left) just as the Sun was setting. This image was taken from 3.5 million kilometres away, before NASA's New Horizons passed by Jupiter on 28 February, on its way to Pluto. On Jupiter's moon Io (left), an eruption from the volcano Tvashtar spews out its dusty plume as seen from 2.5 million kilometres away.

NASA/JOHNS HOPKINS UNIV./SOUTHWEST RES. INST.

SKIMMING THE ▶ SURFACE OF MARS

Rosetta, the European comet-hunting mission, skirted Mars at a distance of just 250 kilometres during its flyby on 25 February, and used the opportunity to test its instruments. This true-colour image was taken the day before the closest approach, from 240,000 kilometres away. Rosetta will now team up with New Horizons to take more detailed measurements of Jupiter, especially the charged gases flying around its moon Io, before it returns for a flyby of Earth.
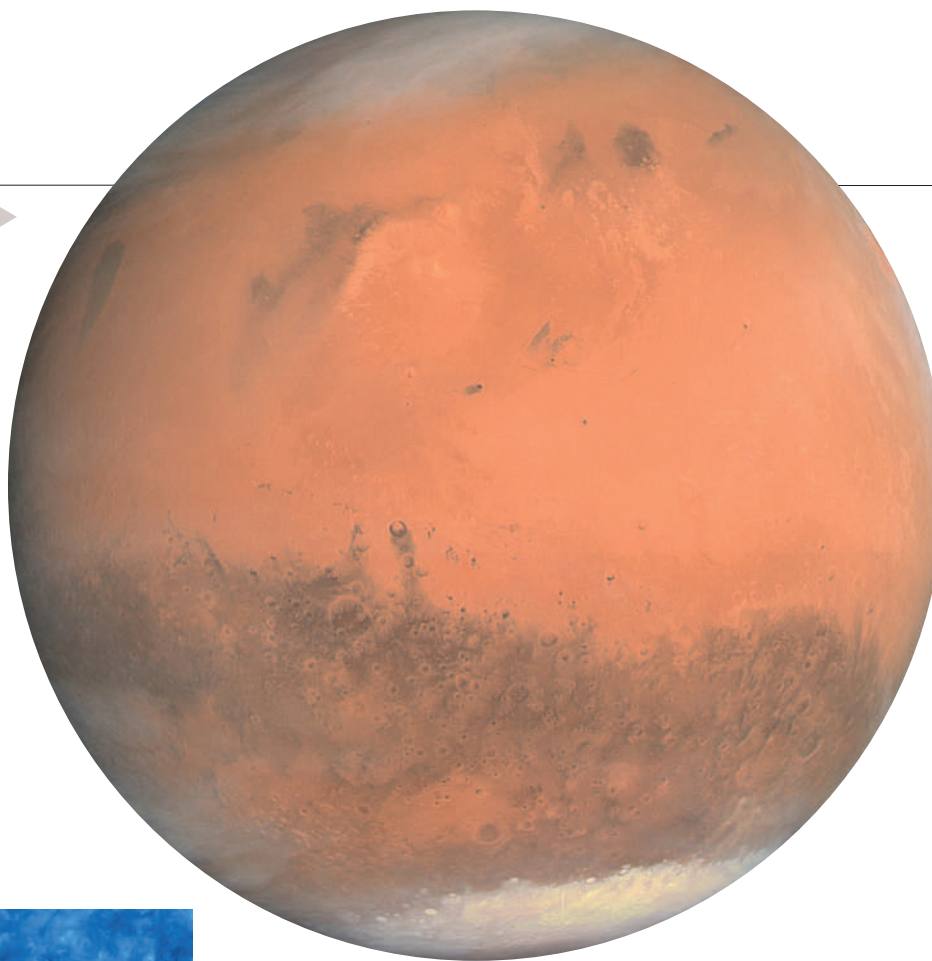
ESA MPS/UPD/LAM/IAA/RSSD/INTA/UPM/ DASP/IDA

NASA/NRL
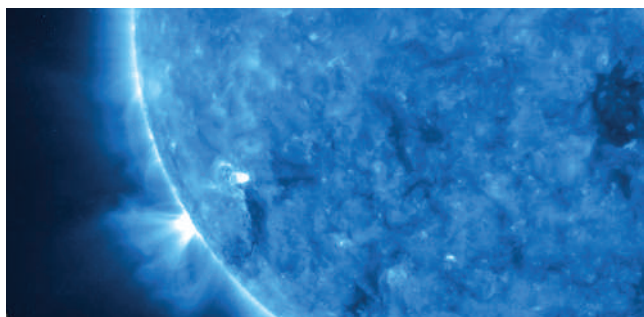
◀EXTREME SUN

The image below is the first ever profile view of the space between the Sun and Earth, captured by NASA's Solar Terrestrial Relations Observatory (STEREO) spacecraft. Compiled from five instruments, the image shows sunlight scattering from zodiacal dust. The white line is Venus, and the blue dots are background stars. Earth would be off to the far left. STEREO also sent back a high-resolution image of the Sun's surface in extreme ultraviolet light (left). Late next month, the mission's two sister craft will start producing three-dimensional views of the inner Solar System, including solar storms travelling towards Earth.

NASA/NRL
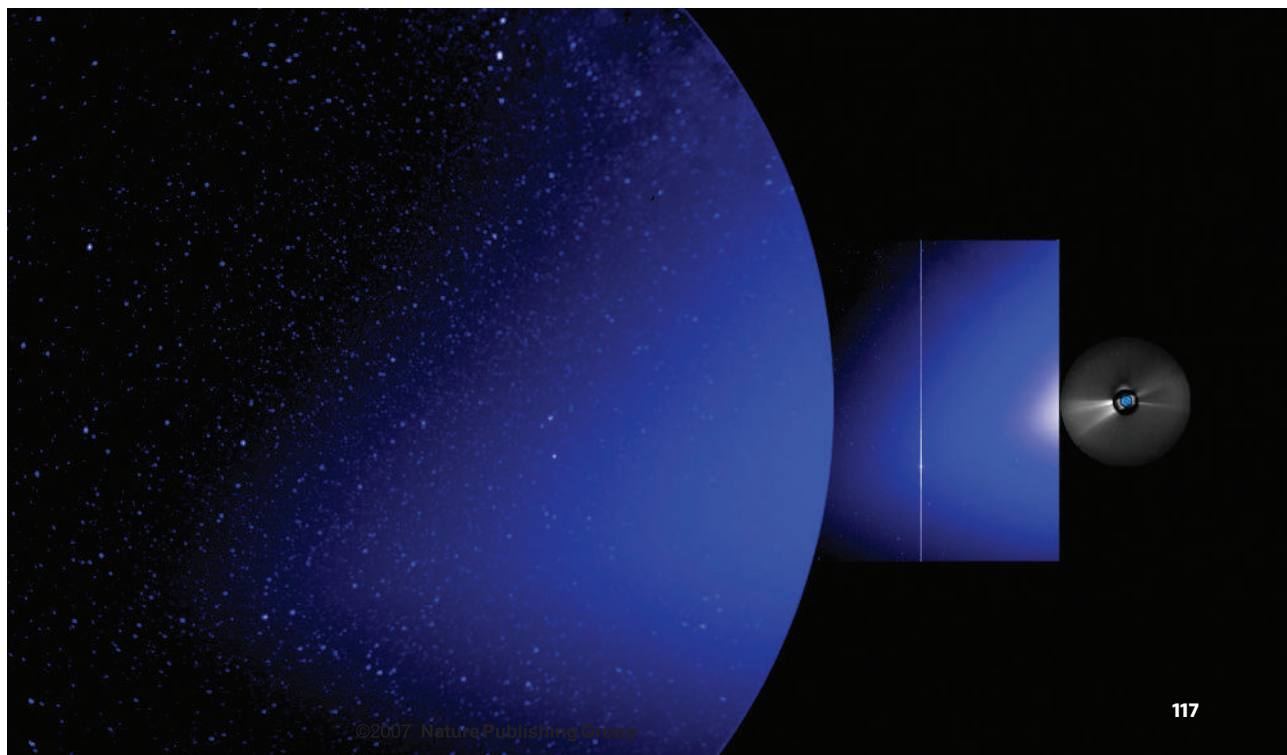
Text by Katharine Sanderson and Geoff Brumfiel.

Biobank provides leads for selecting guide dogs

TOKYO

Biobanks, which combine genetic and lifestyle data from thousands of people, are increasingly being used to help researchers understand the roots of disease. But they can be used to study other things, such as behaviour. An experimental biobank set to open in Japan next month aims to do just that — for dogs.

Four years ago Hiroshi Suzuki, a professor of functional genomics at the Obihiro University of Agriculture and Veterinary Medicine in Hokkaido, found out that Japan suffers from an acute shortage of guide dogs, and decided to do something about it. Training guide dogs is time-consuming and expensive, costing about ¥2.7 million (US$23,000) per dog. And even among labrador retrievers, which are particularly suited to the role, only about 30% of the dogs that are trained work out as guide dogs. Unfortunately, it is difficult to predict which puppies will make good guide dogs, says Takafumi Wada, head of the guide-dog association in Hokkaido, with which Suzuki is collaborating. And that's where Suzuki's database comes in.

Guide dogs have to be neutered before training starts, so successful dogs cannot be used for breeding. Suzuki gets round this by freezing testis and ovary tissue from dogs before training, with a view to breeding from the best. But ultimately he wants to develop genetic tests to select suitable dogs for breeding and to measure puppies' aptitude before they enter training. He adds that a similar approach could be applied to select dogs suitable for police duties or to make good pets.

> "We hope Hiroshi Suzuki will be able to breed more successful guide dogs."

So far, Suzuki has screened DNA from 200 labradors. He has looked at 13 genetic variations called single nucleotide polymorphisms, or SNPs, and found five SNPs in three genes that he believes are related to guide-dog ability. In particular, a gene that codes for an enzyme called COMT (catechol-*O*-methyltransferase), which partly regulates mood in humans, seems to be involved in dogs' ability to concentrate, he says. In Suzuki's study, 80% of dogs with all five SNPs passed their training, whereas the pass rate was only 60% for dogs with just one of the SNPs. Suzuki is now writing up his results for publication.

Suzuki admits that the number of dogs he has studied so far is too small to confirm the role of these genes in determining guide-dog ability, and he is working with guide-dog organizations in Finland and South Korea to collect at least 1,000 blood samples.

"Suzuki has a very innovative approach," says Yoshihisa Yamane, head of the Japan Veterinary Medical Association in Tokyo. "We hope he'll be able to breed more successful guide dogs."

But Masanari Itokawa, who studies schizophrenia at the Tokyo Institute of Psychiatry, is sceptical, pointing out that the link between genetics and behaviour tends to be complex, and that large-scale studies using thousands of SNPs will be needed. "We can't find the relations between gene mutations and personalities or behaviours so easily," he says.

But Suzuki thinks his work, which can draw on the dogs' detailed family histories, could serve as a model for human disease studies. ■

Ichiko Fuyuno

Lab work: a canine biobank aims to ease Japan's shortage of guide dogs.

J. CRAIGMYLE/CORBIS

SCORECARD

▲ **Chinese pest control** Shanghai residents are being encouraged to breed fish that feed on mosquito larvae as an environmentally friendly way to combat this summer's expected plague of the insects.
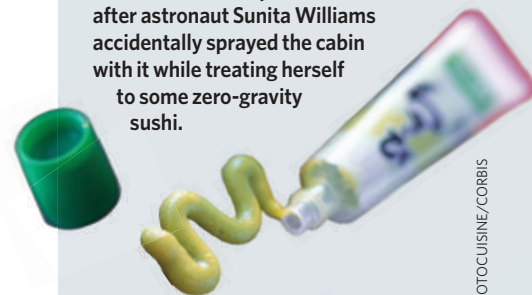
J. SHAW/NPA

▼ **Japanese condiments** A tube of wasabi has been banished from the International Space Station after astronaut Sunita Williams accidentally sprayed the cabin with it while treating herself to some zero-gravity sushi.

BILIC/PHOTOCUISINE/CORBIS

ON THE RECORD

" **There's a trend but the error is big (increase the *N*),**

The rat's no good, use the guinea pig (increase the *N*),

The antibody's weak, the background high (increase the *N*),

My *P* value is bigger than pi (increase the *N*) "

Lyrics to *Increase the N*, a Ramones-esque tribute to science by Canadian rockers HEFE. To hear it, visit www.myspace.com/heferocks

WORDWATCH

Conservapedia: The new conservative encyclopaedia (www.conservapedia.com) that allows conservative Christians to hit back at the 'inherent liberal bias' of Wikipedia. Under 'Global warming', for example, it warns us that "scientists are mostly liberal atheists, untroubled by the hubris that man can destroy the Earth which God gave him".

Sources: Reuters, Associated Press, The World's Fair

SIDELINES

Superconductivity two decades on

DENVER, COLORADO

They were lionized in the press, celebrated by the US government, and even given exclusive access to New York City nightclubs — that was 20 years ago. This week, the physicists who made key early breakthroughs in the field of high-temperature superconductivity recalled their fame and expressed hopes that the field would once again dramatically move forward.

Back then, in a marathon session on the evening of 18 March 1987, at a meeting of the American Physical Society, researchers discussed a new-found class of 'superconducting' material that carried electrical current without resistance at dramatically higher temperatures than its predecessors. The impromptu lectures at the New York City Hilton ran until after 3 a.m. and were later dubbed the "Woodstock" of physics. "The euphoria was really unbelievable," says Paul Grant, a physicist at Stanford University, California, who was there that night.

By contrast, the mood was nostalgic at this year's American Physical Society meeting, held this week in Denver, Colorado, when many of those at the original 1987 session gathered for a reunion and to discuss the future.

Superconducting materials differ markedly from other conductors. Normally, the atomic lattice in a metal causes electrons to scatter, creating heat. But in superconductors, the electrons pair together in a way that allows them to travel effortlessly through the material. Superconductivity was first discovered in 1911, and described in theoretical calculations in 1957 by John Bardeen, Leon Cooper and John Robert Schrieffer.

The materials offer the tantalizing possibility of allowing energy to be transmitted with virtually no loss, promising powerful motors and better generators. But the low temperatures at which superconductors originally worked prevented them from being practical. For most of the 1970s and 1980s, the record for a high-temperature superconductor remained at a miserable 23 kelvin, or 23 degrees above absolute zero.

Then in 1986, Georg Bednorz and Alex Müller of IBM's Zurich Research Laboratory in Rüschlikon discovered a new class of superconductor that worked at 30 K. These cuprates, as they are known, consisted of layers of copper oxide with small amounts of other materials added. The IBM work was soon confirmed by others, including Paul Chu of the University of Houston, Texas, who discovered an

The "Woodstock" of physics: 18 March 1987, when physicists met to discuss a stunning class of material.

AM. INST. PHYS.

yttrium–barium–copper oxide material that could superconduct at 93 K, above the temperature of liquid nitrogen. The discovery meant that superconductors would be cheaper to cool and easier to use. "It created the feeling in many people that nothing is impossible," says Bednorz, who won the 1987 Nobel physics prize with Müller for their discovery.

Condensed-matter physicists found themselves turned into overnight celebrities. "New Superconductors Offer Chance to Do the Impossible," trumpeted *The New York Times*. Physicists were invited to the White House for a demonstration to then-President Ronald Reagan and given free passes to a New York City discothèque. "It was a heady, heady experience," says Grant. Buoyed by the promises, Congress passed the Superconductivity Competitiveness Act of 1988, which boosted funding for the field.

But there remained a problem. Nobody knew exactly how the new materials worked. After their initial discovery, little progress was made in increasing their operating temperatures. And, because the materials were ceramic, they proved difficult to manufacture and incorporate into existing infrastructures.

Today, says Grant, little has come of those early claims. "No high-temperature superconducting technology is turning a profit," he says. "If you reflect back on 1987 there were a lot of futures promised, I think irresponsibly."

And the scientific field has also slowed. Although more than 100 superconducting

> **"It created the feeling in many people that nothing is impossible."**

cuprate materials have been discovered, the record for superconductivity remains at 164 K, about halfway between absolute zero and room temperature. The theoretical understanding of the material is incomplete as well, says physicist Douglas Scalapino of the University of California, Santa Barbara. Although many researchers believe that electrons pairs underlie superconductivity even in these new materials, no one knows how the electrons bind together. "We don't understand what causes it," he says.

A low point came last September when two German researchers predicted that, given current trends, publications of papers in high-temperature superconductivity would end between 2010 and 2015. The researchers later revised their estimates, but not before angering many scientists (see *Nature* **443**, 376–377; 2006).

But despite the frustrating lack of progress, many remain hopeful. Room-temperature superconductors may yet be found, says Chu: "At this moment, there's no evidence, either theoretical or experimental, that tells us it's impossible."

And high-temperature cuprates may soon find some applications in defence, says Grant. The materials' electrical efficiency may make them useful for high-power microwave beams that can be used to disable small ships.

Optimists say the field could be reinvigorated at any time by a new discovery, whether through serendipity or otherwise. "If something happens, there will be another Woodstock and another explosion," Grant says. ∎

Geoff Brumfiel

Design chosen for nuclear warheads

The Lawrence Livermore National Laboratory in California has won a competition to design a new nuclear warhead for the United States. The blueprint could become the basis for the first such weapon produced in more than 20 years.

The Reliable Replacement Warhead (RRW) is intended to replace the 10,000 or so weapons in the current nuclear stockpile, without the need for testing. The United States has maintained a moratorium on the testing of nuclear weapons since 1992.

The design by the Livermore laboratory and Sandia National Laboratories in New Mexico was chosen over that from the Los Alamos National Laboratory because it was based on previously tested components that would, officials say, provide greater confidence that it could be certified without testing.

The RRW is meant to be a safer and more reliable bomb than the warheads in the current stockpile. For example, it won't use toxic beryllium as today's warheads do, and 'insensitive explosives' will replace conventional ones to trigger the nuclear reaction, lowering the risk that the bomb will go off accidentally. "It's not about starting a new arms race," says Steve Henry, deputy assistant to the secretary of defence for nuclear matters. "The new design gives us an opportunity to move on from methods and techniques used during the cold war."

But critics note that the language the agency has used to defend the RRW programme has shifted since discussions began in earnest in 2005.

At first, the main argument in support of the RRW was that the current stockpile might become unreliable if radioactivity leaking from the plutonium pit damaged the surrounding metal. But research released last November showed that the current generation of weapons will last for 85–100 years (see *Nature* **444,** 660-661; 2006). "Now if you read their press statements, it's all about improving safety and security," says Robert Nelson, a physicist at the Union of Concerned Scientists' Global Security Program in Washington DC.

Others argue that if the RRW project goes ahead, it could actually increase pressure to resume testing. One of its goals is to keep generations of engineers and scientists trained in nuclear-weapons skills, which would require designing and redesigning the warheads. Philip Coyle, senior adviser to the Center for Defense Information in Washington DC, says that he finds it hard to see how the cumulative effects of successive modifications could be certified without a return to testing.

But for the programme to move forward, the costs and a production plan will need to be agreed by Congress, some key members of which have already expressed their opposition to the project. ■
Lucy Odling-Smee

> **"The new design gives us an opportunity to move on from methods used during the cold war."**

Approval for cattle drug imminent despite concerns

A new antibiotic for cattle is poised to be approved by the US Food and Drug Administration (FDA) even though it could cripple a potent human antibiotic by promoting the rise of resistance to it.

The cattle drug, cefquinome, would join more than a dozen effective treatments already available for bovine respiratory infections. It belongs to the same family of drugs as cefepime, the only drug in the class approved for human use in the United States. Cefepime is used to treat serious infections in cancer patients.

An FDA advisory committee voted 6–4 against approving cefquinome last September. But agency officials told *The Washington Post* that approval was likely on the basis of an agency rule constraining its ability to block animal drugs on human-health grounds.

Latest flu sample will come from 1918 pandemic victim

The body of a British diplomat who died 88 years ago is be exhumed in a bid to learn more about avian flu.

Mark Sykes died in Paris in 1919 while

Virus particles from the body of flu victim Mark Sykes may shed light on the virus's lethality.

taking part in peace negotiations. He was one of 60 million to die as Spanish flu swept the globe at the end of the First World War. The disease is thought to have been caused by an avian flu virus similar to the H5N1 strain that has infected birds in Asia and Europe, so samples from those who died could help prepare for a future epidemic.

Only five samples are currently available for study from people who died in the epidemic, says John Oxford, a virologist at Queen Mary, University of London. The amount of tissue that can be removed from these samples is limited, so more samples would be a valuable addition to the set, he says. The other samples are also all from the lung, but Oxford plans to take tissue from several of Sykes's organs.

Oxford received permission from religious authorities to exhume the body late last month and hopes to begin work later this year after receiving ethics-committee approval.

Row over Aboriginal remains rumbles on

London's Natural History Museum is to meet with Aboriginal groups in a bid to resolve a legal tussle over human remains held by the museum.

Last November, the museum pledged to return the remains to the Tasmanian Aboriginal Centre, a Hobart-based organization that represents the descendants of the communities from which the remains were taken (see *Nature* 444, 411; 2006). But the museum first wants to run a variety of tests, including computerized tomography scans. Researchers say such scans are important because so little similar material is available for study.

Plans for the samples' return were delayed in early February after the Tasmanian centre asked a British court to block the tests on the grounds that the remains are sacred. The museum now says it will meet with officials from the centre within the next couple of months in

TOPICAL PRESS AGENCY/GETTY IMAGES

a bid to sort out the issue, although the original return date of 31 March is now unlikely to be met.

Animal extremists 'behind Oxford bomb plot'

Animal-rights extremists have been linked to two bombs planted at the University of Oxford, UK, last month. The devices, found at Templeton College on 26 February, did not detonate, and it is unclear whether they would have caused significant damage.

Bite Back, an anti-animal-research website that documents the activities of extremists, has claimed that the British-based Animal Liberation Front planted the bombs. "This latest action is part of an ongoing fight against the University of Oxford and its continued reign of terror over the unseen victims inside its animal labs," a statement on the website says.

The incident seems to mark the latest in a series of confrontations over the university's new biomedical research lab, still under construction. It took place a year after the first street demonstration in favour of animal research at the university (see *Nature* 440, 10–11; 2006).

Peruvian pillars stake claim as ancient solar observatory

Archaeologists have identified the remains of the oldest-known solar observatory in the Americas. A set of 13 'pillars' at Chankillo, Peru, mark out the positions of the Sun throughout the year, and date back to the fourth century BC.

Although the existence of the cuboidal structures (pictured) was already known, their purpose had not been realized until now. Archaeologists Ivan Ghezzi of Yale University in New Haven, Connecticut, and Clive Ruggles of

I. GHEZZI

the University of Leicester, UK, report that, when viewed from one of two observation points, the pillars describe the rising and setting of the Sun throughout the year (I. Ghezzi and C. Ruggles *Science* **315,** 1239–1243; 2007). The structures are some 500 years older than similar monuments of the Maya civilization of Central America, and predate the European conquests by almost two millennia.

US women's health office faces budget hatchet

The US Food and Drug Administration (FDA) is considering a plan to chop a third off the budget of its small office dedicated to women's health.

In a spending plan being developed for the remainder of 2007, FDA administrators have proposed cutting $1.2 million from the $4-million budget of the Office of Women's Health. Andrew von Eschenbach, the FDA commissioner, said that no final decision had been made about the spending plan.

Critics charge that the proposal amounts to retaliation by the Bush administration against the stand taken by the women's office in favour of making Plan B, the 'morning-after' contraceptive, available without prescription (see *Nature* 437, 179; 2005).

BUSINESS

A place in the sun

In the Tuscan countryside just outside Siena, a historic laboratory is helping Novartis in its bid to become the world's premier vaccine company. **Alison Abbott** reports.

No one attending one of Rino Rappuoli's symposia in Siena, Italy, could fail to be seduced by his laboratory's melding of modern efficiency and old-world charm. The Sclavo Institute may have top-of-the-range facilities, but there is still scope for lunch on the terrace accompanied by a glass of chianti and uninterrupted views of the Tuscan hills.

What might surprise some outside visitors enjoying the lab's exceptional ambience is its newfound role as the intellectual engine behind the multibillion-dollar vaccine arm of Swiss drug company Novartis.

"The research group at Siena has a long history of being at the cutting edge," says Stanley Falkow, a molecular microbiologist at Stanford University in California. So when Novartis took control of the lab, as part of its acquisition of California biotech firm Chiron last April, it inherited a broad portfolio of vaccine candidates, he says.

On 28 November, Novartis, satisfied with the fruits of its acquisition to date, declared its intention to become the world leader in vaccines. Last year, it was the fifth-largest supplier in a US$11-billion market, with $965 million in vaccine sales.

The tradition of vaccine research on the Siena site goes back a long way. In 1904, Achille Sclavo, who developed the first protective serum against anthrax, founded a vaccine and serology institute at his home outside the old city walls. Sales of antisera and vaccines against diseases such as anthrax, typhus and cholera supported his research and the growth of the institute. By the 1950s, the company was not only fulfilling all of Italy's vaccine needs, it was also supplying other countries including the United States.

But things later flagged. In 1983, the Sclavo Institute was acquired by the state-owned energy giant Eni. It was privatized in 1990 and then sold to the electronics firm Marcucci, which wanted to close the lab and build a hotel on the land.

Rappuoli — then Sclavo's director of research — rallied scientific friends around the world in support of the institute, and the government in Rome intervened, barring Marcucci from ending research at the site. Rappuoli then helped the company to find a buyer that wanted to exploit the research potential. Chiron, which was trying to develop AIDS vaccines at its headquarters in Emeryville, bought the lab in 1992 for an undisclosed price.

Chiron put Rappuoli in charge of its vaccine development, adding investment but interfering little with his personal style of open interaction with the academic world.

Today, the business has half-a-dozen vaccines in late-stage clinical trials, for diseases including meningitis and influenza, and many more in earlier stages of development.

And Novartis is widely expected to win European regulatory approval within the next few months for the world's first influenza vaccine that can be produced by cell culture, a procedure designed to replace the time consuming egg-based process. According to Novartis officials, the US Food and Drug Administration has indicated that it may not demand fresh US clinical trials if the product — OptaFlu — is approved in Europe.

"The potential of the vaccine pipeline is substantially larger than we anticipated," says Karl Heinz Koch, an analyst who watches Novartis for the Swiss bank Vontobel. "The cell-based flu vaccine alone could win over a significant part of the $6-billion-a-year flu-jab market."

Before it acquired Chiron, Novartis was not active in vaccines. But the company now aims to have doubled last year's sales by 2011, with the main earners being vaccines for influenza and for travel-related disease risks. It has pledged major investment in its vaccine programme, while absorbing Chiron's pharmaceutical division into its own.

According to Falkow, much of this vaccine success comes down to the research at the Siena laboratory — the source of most of Novartis' approved products as well as the half-dozen candidates in trials — and Rappuoli, who is now in charge of Novartis vaccine research worldwide.

About a fifth of the lab's scientists are not Italian — quite a high proportion by Italy's standards. And although employed directly by the drug company, they are an integral part of the wider microbiology community, outsiders say. The staff speak openly about much of their research at meetings, publish in top-ranking journals and are acknowledged leaders in some areas of fundamental research. "Most of the Siena scientists are high level, publishing as much of their work as they can," says Philippe Glaser of the Pasteur Institute in Paris. "This situation is down to Rappuoli — he's an extremely good researcher and a good manager."

Fresh targets

The 150 or so scientists on the Siena site — there are another 100 physicians and technicians working in clinical research — have made several important discoveries in the biology of infectious disease. They found, for example, how the bacterium *Helicobacter pylori* causes the specific damage to the stomach cells it infects that leads to ulcers or cancer.

They are also pioneering what Rappuoli calls 'reverse vaccinology', a technique that uses information from a microbe's genome to predict which of its proteins might be useful for vaccines.

One of the lab's most recent innovations was to use this method to identify structures called 'pili' in *Streptococcus* bacteria, which cause diseases such as meningitis. Pili are thin tubes of protein that protrude from the cell walls of the bacteria, helping them to stick to the mammalian cells they are going to infect. Scientists at Siena are now studying the feasibility of

The Sclavo Institute has produced a treasure trove of vaccine candidates for Novartis.

Rino Rappuoli opts for open interaction with the academic world.

NOVARTIS

NOVARTIS

developing vaccines based on pili proteins.

Rappuoli says he welcomed the takeover by Novartis. Despite Chiron's innovative approach to product development and considerable commercial success, it never managed to bring a blockbuster drug to market, he points out. It just didn't have the cash to push all of its promising vaccine candidates through clinical trials. "Now Novartis is bringing them through fast,

"Most of the Siena scientists are high level, publishing as much of their work as they can." — Philippe Glaser

running trials in parallel," Rappuoli says. And researchers at the bench are pleased to have access to some of Novartis' powerful research resources in areas such as bioinformatics.

The future may be less bright for the 70 or so vaccine researchers who worked at Chiron's other main laboratory at Emeryville. Rappuoli is moving the unit to Boston, to be close to the drug researchers at the Novartis Institutes for Biomedical Research. "In the next 15 years," he says, "we'll pioneer a new field of combined therapy and immunotherapy." Combined vaccine and drug approaches could yield dividends in areas such as cancer or AIDS, he adds.

Novartis is already bankrolling another of Rappuoli's long-term visions — a separate research institute on the Siena campus, dedicated to developing vaccines for use in developing countries. Novartis has agreed to equip and staff the laboratory, which will look to foundations such as the local Monte dei Paschi di Siena and the Bill & Melinda Gates Foundation to support its research projects. ∎

IN BRIEF

FLU VACCINE The US Food and Drug Administration is set to approve the first vaccine for use in an avian flu pandemic. An advisory committee said the regulator should approve the vaccine, made by Sanofi-Pasteur, the vaccine arm of Paris-based Sanofi-Aventis. The vaccine against the H5N1 strain of bird flu is only mildly effective, however, and will not be sold commercially. It is being bought by the US government for limited use in the early stages of a pandemic, before a vaccine better matched to the actual pandemic strain becomes available (see *Nature* 437, 619; 2005).

GREEN BUYOUT Plans for eight coal-fired power plants in Texas look likely to be scrapped as part of a proposed buyout of Dallas-based electricity generator TXU Energy. The US$45-billion deal between TXU and investors led by Kohlberg Kravis Roberts & Co. and Texas Pacific Group was endorsed by major environmental groups when it was announced on 26 February. The buyers say that as well as revising the plans for new power stations, they will commit to cut carbon dioxide emissions back to 1990 levels by 2020, and adopt strict environmental rules.

ABANDONED SPACE A six-year effort to launch a commercial space facility in Texas has been abandoned. Brazoria County, outside Houston, created the Gulf Coast Regional Spaceport Development Corporation in 2000, hoping that it could find a private company to develop leased land into the world's first base dedicated to commercial space travel. But the county government said on 27 February that it was giving up on the idea.

MARKET WATCH

CLEAN-ENERGY STOCKS
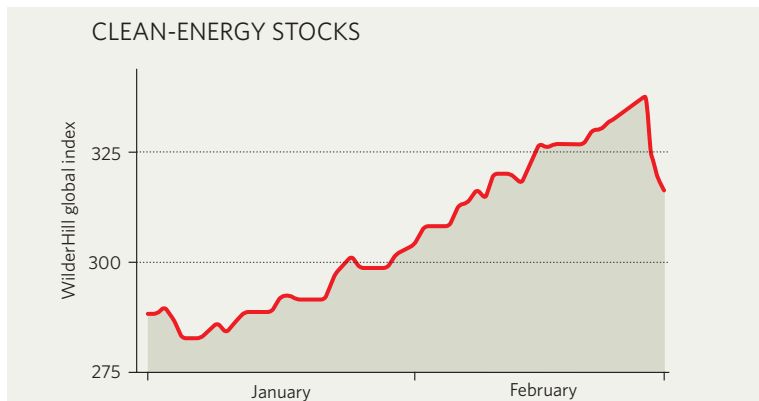
The stars have certainly been aligned in favour of clean-energy stocks early this year, as growing acceptance of the need to reduce greenhouse-gas emissions has bolstered the value of companies with interests in non-fossil-fuel sources of energy.

For 18 months, *Nature* has been tracking the value of a group of US-based energy companies. But this week we shift attention to an international selection of companies, which reflects the market for clean energy outside the United States. The WilderHill New Energy Global Innovation Index — NEX on the American Stock Exchange — tracks the weighted values of about 60 of these companies around the world.

At least until global markets dipped last week, the NEX index had been moving strongly upwards. The release of the fourth report from the Intergovernmental Panel on Climate Change on 2 February lent further impetus to a corporate trend in favour

of clean energy, says Rob Wilder, founder of WilderShares, the California company that co-compiles the index with London-based New Energy Finance. "Even conservative people have been pushing their reservations aside and getting on board," he says. "They are saying: 'we don't love Al Gore, but there's money to be made'."

And Wilder notes that quiet and determined corporate moves into the sector — such as a low-key 8 January announcement by US food giant Cargill that it is creating a subsidiary to build ethanol-processing plants — hint at good long-term prospects.

Even the global stock market correction on 27 February, which saw the Dow Jones fall by more than 400 points, confirmed the ebullience of the sector, says Michael Liebreich, head of New Energy Finance. "It was obviously hit by the correction," he says, "but less so than some main market indices." ∎
Colin Macilwain

BURIED TREASURE

Air bubbles trapped in the Antarctic ice sheet could yield precious information about Earth's climate more than a million years ago. But to access this record, scientists first have to climb one of the coldest peaks on Earth. **Nicola Jones** reports.

Dome Argus in Antarctica is a silent and lonely place. Snow stretches to the horizon in all directions, unbroken by any sign of life. The mountain is one of the least accessible parts of the frozen continent, and quite likely the coldest place on Earth — although no one has been there long enough to catch a record-breaking low. People first stepped onto the summit just two years ago, after a Chinese team made an arduous month-long, 1,200-kilometre trek from the coastal station of Zhongshan.

If you think that buried treasure should be hidden somewhere as remote as possible, then Dome Argus, or 'Dome A', would be an ideal spot. But if you think that it should be gold and shiny, you will be disappointed. Dome A's bounty, if it exists, is tiny pockets of gas trapped in its depths.

Antarctica's great ice cap is crowned by several flat snowy summits known as Domes A, C and F (see map). Reaching the top of Dome A, the tallest and the last to be conquered, was hailed as a triumph of exploration. The scientific importance, though, is not the height itself (although astronomers are excited by the clear air that comes with it) but the ice below. From more than 3,000 metres below Dome A's 4,093-metre peak, researchers hope to extract the oldest ice core in the world, and with it a treasury of climate information.

As part of the International Polar Year, which launches this month, a Chinese team plans to return to Dome A this austral summer to set up a camp. Next year, a larger, international group intends to storm the dome in aircraft to map the ice below. And the Chinese will return too, this time lugging a giant drill. Dome A might not be so lonely or so quiet for long.

An ice core from Dome A would join a formidable list of samples taken from previous Antarctic expeditions. A core from the Russian station Vostok eventually reached 3,600 metres deep and yielded measurable ice some 420,000 years old[1]. A core from Dome C reached only 3,200 metres down[2] but, thanks to a better preserved bottom section, pushed 800,000 years back in time. And a 3,000-metre core being drilled at Dome F caused fleeting excitement last year when the Japanese team spearheading the project said that the ice might be even older.
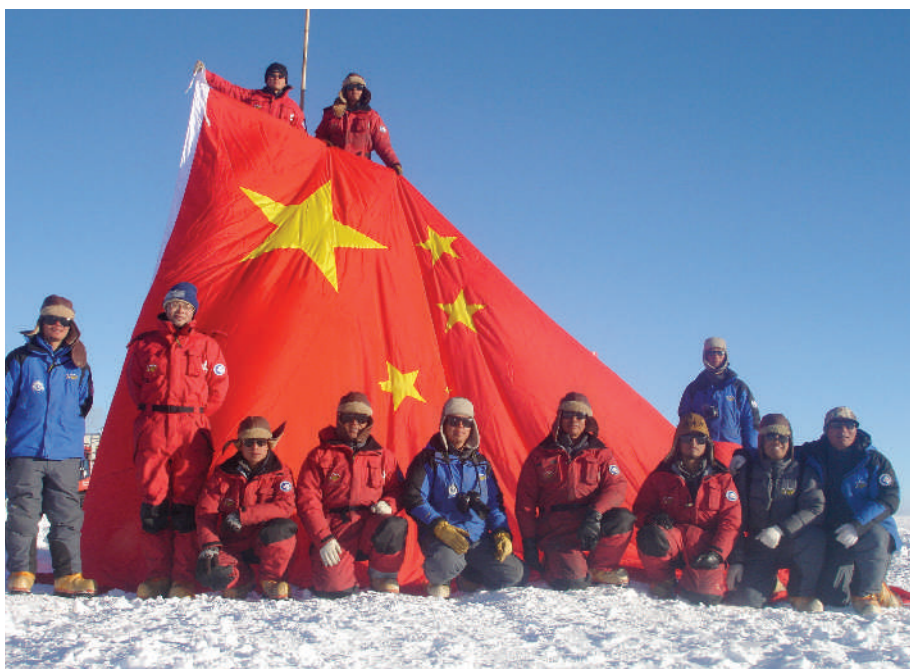
A Chinese team hopes that Dome A will provide the oldest ice core in the world.

But more recent tests have shown that the usable ice from close to the bottom of the dome was some 720,000 years old, and the team now suspects that the last few chunks will hold only a few thousand years more.

A wealth of information

The reason for this old-ice treasure hunt is simple. Earth used to experience periods of glaciation once every 41,000 years or so. Today, glacial cycles come every 100,000 years. Evidence from sediment cores suggest that the key transition between these states took place over a period of several hundred thousand years, about a million years ago. No one knows why it happened. One idea is that levels of carbon dioxide in the atmosphere plummeted and cooled Earth enough for a substantial extra layer of ice to form. This massive burden of ice would have made it hard for the planet to respond so nimbly to the orbital drivers of climate change, thus shifting it into a more stately pace of glaciations.

The only way to confirm the idea is to find air bubbles that date back 1.5 million years or so, and track how the carbon dioxide levels changed over time. Ice is the only place to find such bubbles. And for old ice, Antarctica is the place to go. Ice cores from Greenland, where the ice flows more dynamically, reach back only 100,000 years or so. But Antarctica has been covered in ice for an estimated 30 million years, and models of glacial flow suggest that there could be an uninterrupted record of ice that stretches back a few million years — probably beneath Dome A. The International Partnership in Ice Core Science (IPICS), a 19-nation group co-chaired by Eric Wolff, an ice-core specialist with the British Antarctic Survey in Cambridge, UK, and Ed Brook of Oregon State University in Corvallis, has made finding that ice one of its main goals.

Dome A not only has plenty of ice to drill, it also gets very little snow — just 1.25 to 1.5 centimetres of its equivalent in water per year, compared with the 3 centimetres at Domes C and F or the 50 centimetres that dumps on the coastal station of Halley. That means that the ice on Dome A contains snow from a very long time period.

But there are complications. The Antarctic ice cap acts as an insulator, lying like a blanket across the continent and trapping geothermal heat below it. The thicker the ice, the greater the insulation, and so when the ice gets really thick its base will frequently become warm enough to melt, shortening the record.

Also, the topography of the rock beneath the ice is complex, to say the least. A mountain range lies down there (see 'The hidden mountains', overleaf). If the ice at the bottom of the ice sheet has been forced to move up and over rocky ridges, it will be folded, muddled and mixed, making it impossible to date it or to extract clean information from it. The bottom 70 metres of the Dome C core were like this, making its oldest ice unusable.

Models suggest that the flow of ice away from the base of Dome A is small, so older ice should still be preserved at the bottom. A map of ice ages by modeller Philippe Huybrechts of the Dutch-speaking Free University Brussels in Belgium (see map), confirms that an area near the peak of Dome A — a vast swath about the size of Britain — is suitable for an old-ice hunt.

Apart from Dome A, other candidate sites for the oldest ice do exist. The Aurora basin, near Dome C but closer to the coast, for example, is about 4,500 metres deep and could potentially hold very old ice. Australia plans to drill a 400-metre test core there in the 2008–09 season, says Vin Morgan of the Australian Government Antarctic Division near Hobart, Tasmania. But this area is lower and warmer than Dome A, increasing the chances that its bottom ice has melted substantially. The nearby Astrolabe basin has an even

deeper 4,700 metres of ice, but covers a small area in which the surrounding rock may have distorted the ice at the bottom. Farther afield, there are other areas in Antarctica likely to hold very old ice (see map), but they are trapped in mountain ranges where the ice record is much thinner and more squashed.

Disturbing the peace

Dome A thus remains the prime candidate for drilling. But too little is known for researchers to draw an X in the snow and plant their drill. So starting next year, if funding comes through, the sky over Dome A will be filled with the rare noise of low-flying planes, burdened with radar and equipment to measure gravity and magnetic fields.

Particularly helpful will be the radar surveys, which can pick up changes in density, crystal structure or dust content in the ice. By flying from Dome C to Dome A, the IPICS team hopes to be able to track ancient layers along the 1,000-kilometre flight path, thus revealing the depth of correlated layers in the ice at Dome A.

Given the remoteness of Dome A, the planes will need local bases from which to refuel. Ideally, these will be placed around the dome at slightly lower altitudes, as the height of the dome makes the air so thin that propeller planes have trouble taking off and pilots' functioning can be impaired. "Technically, the pilots should be using supplementary oxygen," says Wolff. And in similar circumstances, loaded Twin Otter planes have had to use jets strapped to their wings to gain enough lift.

That same season, the Chinese researchers plan to return, this time carrying French

POLAR YEAR PROJECTS

The fourth International Polar Year (IPY) is a bit of a misnomer — it's actually two years long. From March 2007 to March 2009, a host of scientists will head out to both the Arctic and Antarctic for targeted research, from marine biology to anthropology. Here's a look at some of the projects being planned — although funding for some is still pending.

Whale movement

Several hundred beluga whales (*Delphinapterus leucas*) in the Arctic will be tagged with satellite transmitters and, in some cases, oceanographic data collectors. Understanding the timing and pattern of beluga movements in relation to ice and ocean conditions may help efforts to protect the whales in the face of climate change.

Ocean microbes

A Norwegian-led effort to document the biodiversity of microorganisms in the polar seas could provide a basis for understanding how these creatures help to regulate the ecosphere.

Polar astronomy

Because of their extremely cold, dry, stable air, the polar plateaus provide the best sites on Earth for a range of astronomical observations. An Australian-led team will assess just how good the conditions are for astronomy at sites including Dome A on the Antarctic plateau.

Spider survey

German researchers hope to lead a survey of spider biodiversity across the Arctic. Because spiders adjust their lifecycles to microclimatic conditions, studying them can help scientists track the effects of rising temperatures on terrestrial habitats.
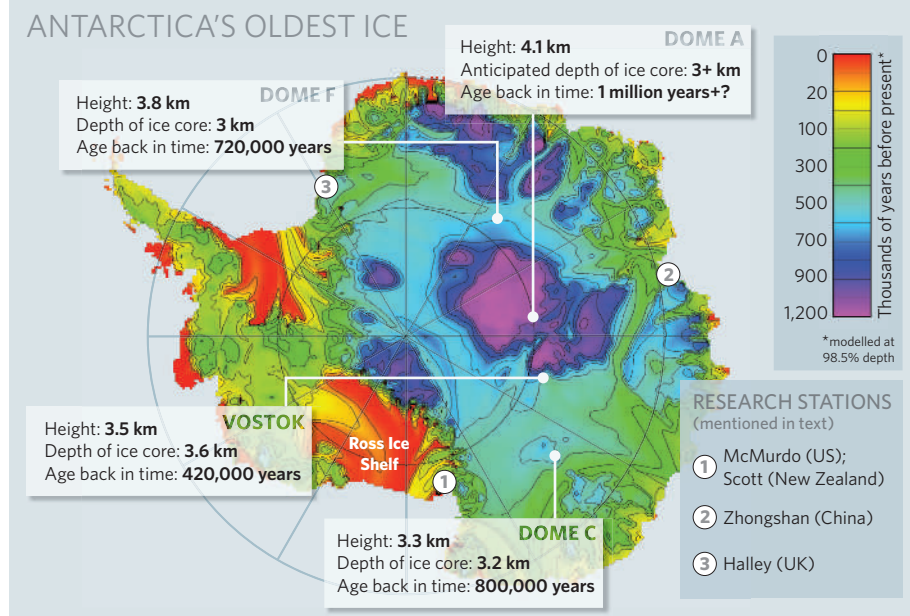
Carbon pools

An international effort, led by Sweden, plans to assess the quantity and quality of organic matter in high-latitude soils. The work may prove crucial to predicting what could happen to the enormous stock of carbon trapped there if the soils thaw.

ANTARCTICA'S OLDEST ICE

DOME A
Height: **4.1 km**
Anticipated depth of ice core: **3+ km**
Age back in time: **1 million years+?**

DOME F
Height: **3.8 km**
Depth of ice core: **3 km**
Age back in time: **720,000 years**

Thousands of years before present*
0
20
100
300
500
700
900
1,200

*modelled at 98.5% depth

VOSTOK
Height: **3.5 km**
Depth of ice core: **3.6 km**
Age back in time: **420,000 years**

Ross Ice Shelf

DOME C
Height: **3.3 km**
Depth of ice core: **3.2 km**
Age back in time: **800,000 years**

RESEARCH STATIONS
(mentioned in text)

① McMurdo (US); Scott (New Zealand)
② Zhongshan (China)
③ Halley (UK)

The hidden mountains

There is a mountain range on our planet as large as the Alps, but that was only discovered in the late 1950s. The reason for its elusiveness? It's buried under Antarctic ice.

Radar surveys of the central east Antarctic, along with data from the gravity-sensing GRACE satellite, have revealed this mystery range beneath Dome Argus (see picture, right, from GRACE radar). The highest peaks look to measure 5 kilometres from base to top. The tallest exposed mountain on the continent is Vinson Massif, at 4,892 metres above sea level.

But data are scarce, and no one knows exactly what these Gamburtsev subglacial mountains look like, or the true size of the range. More importantly, no one knows how they got there. The

BEDMAP CONSORTIUM
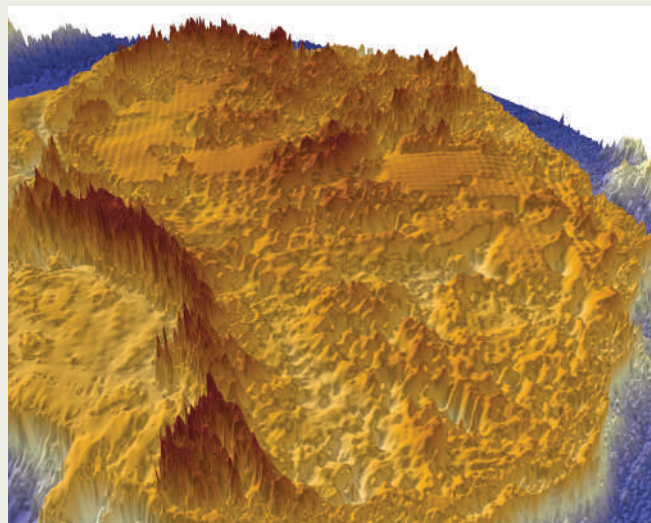
peaks soar upwards from the centre of an old tectonic plate, which geologically ought to be quiet. "It's like finding the Rockies in the middle of the Canadian shield," says geophysicist Fausto Ferraccioli of the British Antarctic Survey in Cambridge, UK.

One idea for how they formed is that a collision of two continental plates more than a billion years ago forced up the rock, which then did not erode thanks to a very hard 'cap' rock. But other theories call into question the assumption that the eastern part of Antarctica is much more stable and sedate than the western region. Perhaps there was, or still is, a hotspot rising in a plume from the mantle beneath. "That would mean volcanism, which would change our view of the east Antarctic forever," says Ferraccioli.

The radar, gravity and magnetism surveys being planned for the region should reveal more details about the Gamburtsev mountains. And so, researchers in Antarctica could uncover not only old ice but also details about the continent's geological history. **N.J.**

---

drilling equipment capable of bagging a 500-metre core — an upgrade on the 110 metres they pulled from the ice in 2005. Their work will be dangerous; last time, a team member fell into a crevasse while working at the summit camp. "He was lucky — it wasn't so deep," says expedition member Shugui Hou, of the Chinese Academy of Sciences' Cold and Arid Regions Environmental and Engineering Institute in Lanzhou. The good news is that summer temperatures are only −35 °C. "When the weather was good we were wearing only one sweater, perhaps because it's so dry," says Hou. "It's quite comfy."

Pooling resources

Funding a serious drilling effort will be a bigger project than any country can tackle alone. "No one nation has the resources. We're going to have to do it internationally," says Alan Rodger, head of science programmes at the British Antarctic Survey. With the help of many countries, Wolff estimates, a site could be selected and a drilling operation under way by 2012.

"We want to drill two cores, to give ourselves two chances," says Wolff. One idea would be to blast through the top part of the first hole with a heat drill, not bothering to pull up usable cores until reaching a predetermined depth. Meanwhile, a second site, tens or hundreds of kilometres away, could be drilled in detail from the surface. This strategy of extracting two cores would be a new one; most other sites poured all their resources into a single drill hole (although

jammed drills have often meant that teams had to start a new core).

The technology, at least, exists. Drilling deep into the Antarctic is tricky but doable. The difficulties lie in such things as getting the drill fluid right. The cold, dense liquid that drillers insert into the hole to stop it from closing up on itself must not dissolve the snow around it or get clogged with ice chips that can form a sludge; kerosene with chemical additives is often used. And when the drill gets near the bottom, where the ice can be near the melting point, the problem is to stop ice that melts and refreezes from jamming the drill bit, by using some antifreeze.

Once the core has been extracted, researchers will use all the techniques they can to date the ice and pull information from it. Old layers of ash from known volcanic eruptions act as date markers, and the top part of the core can be matched up to previous ones, already dated, to pin down the age. Another dating trick for the older ice will be to search for the higher amounts of beryllium produced by the increased flux of cosmic rays when Earth's magnetic field reverses itself[3] — as happened 780,000 and 900,000 years ago. Analysing the amount of nitrogen and oxygen in the ice can also help; the ratio of the two changes in concert with a 23,000-year cycle of alterations in the amount of solar radiation reaching the Antarctic[4].

Technicians will then crush the ice in a vacuum to release the air, and measure the amounts

of greenhouse gases such as carbon dioxide and methane. The ratio of oxygen isotopes can also be used to estimate past temperature. But the big prize will be the carbon dioxide — the amount of the gas in each bubble could confirm the idea that a drop in carbon dioxide caused the change in glaciation cycles a million years ago.

As always, there is still the chance that the ice won't hold the expected treasure. If so, the answer to this million-year-old puzzle may lie elsewhere: perhaps, some speculate, the vast plateau of the Canadian shield was at some point scrubbed clean of lubricating mud by all the glaciers, and it was this, rather than a drop in carbon dioxide, that allowed ice sheets to build up enough to slow the planet's cycle of glaciations. Finding the oldest ice could go a long way to answering these questions.

If not the bounty of plummeting carbon dioxide, then a core from Dome A would still guarantee the gem of an ancient climate record. Xiao Cunde, another member of the Chinese expedition, puts the chances of finding million-year-old ice under Dome A at 95%, and 80% for ice older than 1.5 million years. With odds like that, most would be happy to go on a treasure hunt. ■

Nicola Jones is *Nature*'s online news editor.

1. Petit, J. R. *et al. Nature* **387,** 359–360 (1997).
2. EPICA community members *Nature* **429,** 623–628 (2004).
3. Raisbeck, G. M., Yiou F., Bourles D. & Kent D. V. *Nature* **315,** 315–317 (1985).
4. Bender, M. *Earth Planet. Sci. Lett.* **204,** 275–289 (2002).

See Editorial, page 109.
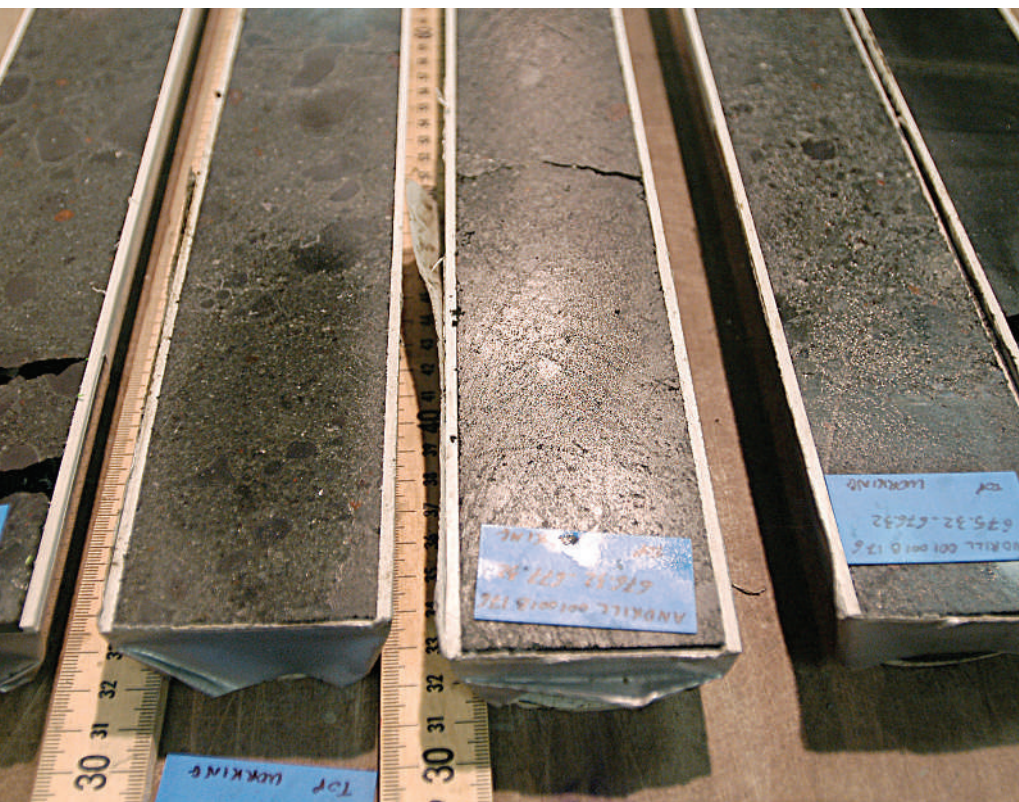
P. WEST/NSF

School of rock

The rocks of Antarctica are obscured literally, and sometimes scientifically, by its ice. But drilling efforts are now showing what we can learn from the hard stuff. **Alexandra Witze** reports.

When it comes to Antarctica's history, ice cores get all the glory. Large-scale ice-drilling efforts, such as Europe's EPICA and Russia's Vostok cores, capture headlines and the lion's share of people and funding. After all, these cores contain air bubbles that are hundreds of thousands of years old, a frozen time capsule from Earth's icy past.

But buried beneath the thick layers of ice, the rocks of Antarctica have far older stories to tell. Trapped within layers of mud and sand are geological records stretching back millions of years. As Antarctica's ice teams continue to hunt for the oldest ice their drills will reach (see page 126), a smaller band of rockhounds is on a similar quest to plug the gaps in the geological record.

The team now has a core that promises fresh insight into how Antarctica's ice waxed and waned over the past few million years. On 26 December, a US$30-million international project called ANDRILL pulled up the final piece of a core from beneath the Ross ice shelf (see map, page 127). Previous coring efforts have offered peeks into Antarctica's deep history — back as far as 34 million years when the continent was first covered in ice. But the new core fills a gap in the ice shelf's history, and sets a new Antarctic record for drilling depth.

The period covered by the core — from the present to more than 5 million years ago — seems to be quite active. Preliminary analysis has revealed thick layers of a greenish rock interspersed throughout the core. This is an indication of open-water conditions, suggesting that the Ross shelf retreated and then advanced at least 50 times within the past 5 million years. With this nearly unbroken record, scientists can explore the history of the shelf in unprecedented detail.

"It's going to be a benchmark that we hope we can refer to for years to come," says Ross Powell, a geologist at Northern Illinois University in DeKalb. "It may be the geological equivalent of the Vostok ice-core record." As project co-leader, Powell is understandably enthusiastic about the core, but then so are other geologists.

The Ross ice shelf, the largest in the world, is a floating extension of the even more massive West Antarctic Ice Sheet. That part of Antarctica is regarded as the most unstable and potentially prone to collapse in a globally warmed world. If the west Antarctic ice sheet melted entirely, it would raise the global sea level by about 5 metres. The Ross shelf is a tiny but important fraction of that. Its behaviour over

Martian mimics

CORBY WEST/JPL

The Phoenix spacecraft will land on Mars during the IPY. Scientists hope to compare data from the planet's northern polar region with soil measurements from an analogous 'extreme environment': the Antarctic Dry Valleys.

Antarctic aliens

More than 40,000 people visit the Antarctic each year — and they probably bring with them the seeds and spores of non-native species. The IPY will see the first full assessment of the environmental impact of these visitors.

Past warming

One way to predict what might happen in a warmer climate is to look back in time. The WARMPAST project will use Arctic sediment cores to reconstruct ocean temperatures for the past tens of thousands of years.

Icy lakes

LANDSAT

Scientists from Canada and Russia will pore over historical data to see how the timing of ice freezing and breaking up on Arctic lakes has changed over the past 50 years.
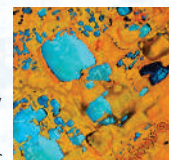
Changes in sea level

UK and other researchers plan to add gauges to monitor the sea level and tides in the waters around Antarctica.

Arctic greening

Climate change is likely to alter the distribution and type of plants at high latitudes. Assessing satellite data, and doing field studies to produce new vegetation maps for Russia, Alaska and Canada, will help scientists predict such changes.

Reindeer herders

The knowledge and experience of nomadic reindeer herders, accumulated over generations, will be documented to establish how herders living across Norway, Russia and Alaska can sustain their way of life.

INGER MARIE GAUP EIRA/EALÁT

Chilled out: the ANDRILL rig on the Ross ice shelf in Antarctica.

day. And stronger-than-expected ocean currents bent the pipe surrounding the drill as it stretched more than 900 metres through the ice shelf and the water below, before finally entering the sea floor. But after two months of non-stop drilling, the team recovered 1,285 metres of rock.

The core, now stored in the freezers of the Antarctic geologic repository at Florida State University in Tallahassee, has yet to reveal all its secrets. The first challenge, says Powell, is pinning down its age. The top half of the core seems to cover the past 5 million years, and that's what contains the 50-plus cycles of ice-shelf collapse. The sediments show repeated layers of ground-up rock debris scraped off the continent by glaciers, interspersed with the open-water greenish ooze rich in the marine organisms known as diatoms. The transitions between the glacial sediments and the open-water ooze seem to be quite sharp, says Powell — suggesting that they took place relatively quickly.

Debris and ooze

Peter Barrett, a geologist at the Victoria University of Wellington in New Zealand, who was chief scientist for the Cape Roberts Project, says he was struck by the number of transitions over such a short period. "I think it's spectacular," he says of the core. "We'd talked about what they might find beforehand of course. But I was surprised at the striking differences that the core brought out."

Another major question is where the ground-up glacial deposits in the ANDRILL core came from. Studying the sediments in the core could help the team trace the material

> **"It's spectacular. I was surprised at the striking differences that the core brought out."**
> — Peter Barrett

— and therefore the ice — back to its source. They might turn out to have come from somewhere else in west Antarctica, says Powell, or perhaps even as far as east Antarctica. Knowing the path the ice took could help researchers better understand how ice flows across the continent, thus aiding future models of ice flow as Antarctica warms.

Other insights could come from comparing the ANDRILL core with ice cores such as EPICA. There isn't much overlap; only the upper 80 metres or so of the ANDRILL rock represent the past 1 million years, and the whole of EPICA covers only the past 800,000 years. But comparing rock and ice records could help palaeoclimatologists correlate increases and decreases in carbon dioxide levels with what the ocean and the Ross shelf were doing at the time. "It will be very exciting to see how the records of ice-sheet changes they have are related to the changes in Antarctic temperature we have," says Eric Wolff, an ice-core specialist with the British Antarctic Survey in Cambridge, UK, and a member of the EPICA team.

The next chapter in the story will start in October, when the second leg of ANDRILL gets under way. For that, the researchers will drill through the 7-metre-thick sea ice in the southern part of McMurdo Sound, through the ocean and again into the sea floor. There, they are expected to pick up where the current core left off — probably around 7 million years ago — and extend the record to around 17 million years ago. With that, geologists hope, they will finally have a complete history of the Ross shelf.

This is key because it seems likely that Antarctica will undergo some serious changes in the future. Both atmospheric carbon dioxide levels and temperatures are projected to increase beyond historical highs over the next several centuries. "Just how long will it be before the temperature increase catches up and we watch the Ross ice shelf go away?" asks Barrett. With the Ross gone, the scenario goes, melting on the main part of the west ice sheet could accelerate. That's something neither the ice experts nor the rockhounds want to contemplate. ■

**Alexandra Witze, *Nature*'s chief of correspondents for America, has been to the North Pole but not Antarctica — yet.**

Warm vents
Studying hydrothermal vents on the Arctic mid-ocean ridge is a challenge as major portions of the ridge lie more than 4,000 metres under pack ice. In July and August, robotic vehicles will dive to the Gakkel Ridge to search for vents.

Satellite shots
The photographic power of Earth-observing satellites is being pooled to yield a wide range of snapshots of the world's polar regions in the highest resolution possible.
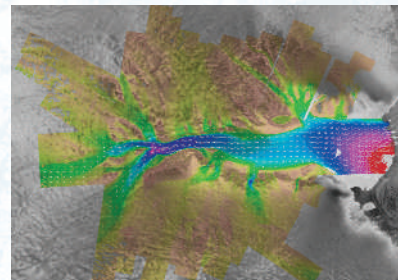
Solar activity
Scientists will take measurements from the polar regions to assess whether variation in the Sun's activity affects Earth's weather and climate by influencing a global electrical circuit in the atmosphere.

Particle physics
Physicists will use the IceCube observatory being built at the South Pole to search for subatomic particles called neutrinos. During the IPY, glaciologists are being invited to use the detectors to study ice flow.

Polar bears
A Danish-led team will examine contaminants in the muscles and bones of polar bears killed by Inuit hunters. Chemical analyses of the bears' body tissues could also shed light on how much climate change is stressing the animals.

Open leads
Some 200 researchers from 15 countries will study the circumpolar flaw lead — an area of open water that forms each autumn when the main Arctic pack ice pulls away from coast. With thin ice becoming more common in the Arctic seas, this region offers a glimpse of how changes to the ice affect ocean life.

Lucy Odling-Smee

Cutting edge: the ANDRILL team recovers the final part of the 1,285-metre rock core from Antarctica.

past millennia could help researchers improve their understanding of how it might respond as temperatures rise in the future.

Conditions on Earth are returning to a state that hasn't existed for millions of years, says David Harwood, a geologist and ANDRILL scientist at the University of Nebraska in Lincoln. "We have to go back to previous times in Antarctica when things were very warm, when carbon dioxide levels were higher, for an analogue of where we're heading," he says.

The history of Antarctica's ice starts about 35 million years ago, when atmospheric carbon dioxide levels and temperatures began to drop.

As the globe cooled, the great ice sheet in east Antarctica began to form. Some time afterwards — perhaps as early as 30 million years ago, or as late as 5 million years ago — west Antarctica gained its ice as well. ANDRILL scientists hope to refine these timing estimates.

Until ANDRILL, Antarctic rock cores have only offered relatively short glimpses of the continent's history. Inland, researchers drilled shallow cores during the 1970s for the Dry Valleys Drilling Project — the deepest reached 300 metres. Offshore, ocean-drilling ships have collected sediment records that were scraped off the continent by the flowing ice; such debris

provides clues to what Antarctica's bedrock looked like long ago. And in the broken-up sea ice that fringes the continent, icebreakers have occasionally been able to grab a quick core of similar sediments (see 'Quick-hit drilling').

Technologically, it is much harder to drill through the thin layer of sea ice that surrounds the continent, or beneath the thicker ice shelves such as the Ross. In the late 1990s, the international Cape Roberts Project tried this for the first time, drilling through sea ice into the sea floor. A severe storm cut one of the seasons short, but the team managed to collect short cores dated to between 34 million and 17 million years ago (T. R. Naish *et al. Nature* **413,** 719–723; 2001). That took scientists back to the formation of the east Antarctic ice sheet, but couldn't tell them what had happened more recently.

Shelf help

ANDRILL aims to fill in that gap. "No one has ever tried to recover a long sedimentary record from under an ice shelf," says the project's other co-leader, Tim Naish of the Institute of Geological and Nuclear Sciences in Lower Hutt, New Zealand. The 200-plus consortium of researchers, run by the United States, New Zealand, Germany and Italy, has allocated two field seasons for drilling two separate cores: the one just completed through shelf ice, and another to be drilled through thinner sea ice.

For this season's work, which began in November near the New Zealand Scott base, the drilling company had to devise a hot-water system to get through the 85 metres of ice shelf, and keep it from refreezing around the pipe. The solution was a giant metal 'doughnut' filled with hot water that continuously ran up and down the drill shaft, keeping the surrounding ice melted enough for operations.

Drilling was tricky at first. Tides cause the ice shelf to flex up and down by about a metre per cycle, and sideways by about half a metre per

P. WEST/NSF

Quick-hit drilling

There are obvious advantages to using a drill rig and drilling for as long as you can — a very long continuous core, for instance.

But there are lots of places in the Antarctic where drill rigs just aren't feasible, either as fixed platforms such as ANDRILL or onboard ocean-drilling vessels that work offshore. In particular, the broken-up sea ice around the edges of the continent makes it difficult for core-seekers to get to where they want to go. They may know exactly

where the sediments they seek are, but they can't get to the spot because of dangerous ice floes.

Overcoming that problem is the point of the SHALDRIL drilling platform. It uses icebreakers to swoop into the region of interest, fending off sea ice as they go, and retrieves short cores from the sea floor. "The object is to give you a lot more mobility," says John Anderson, a geologist at Rice University in Houston, Texas, and co-leader of the project.

But the 'drill and run' strategy hasn't always worked. The first SHALDRIL cruise, in 2005, aboard the US icebreaker *Nathaniel B. Palmer*, successfully acquired cores at sites along the Antarctic peninsula. But the following year, the second cruise ran into major problems with sea ice (J. B. Anderson *et al. Eos* **87,** 402, 408; 2006). Thick ice floes battered the icebreaker, making it impossible to retrieve the planned cores. Instead, the team grabbed short

samples quickly from various locations and left.

Although the team currently doesn't have funding for a third cruise — or even to study the cores it has already retrieved — Anderson says he hasn't given up on SHALDRIL as an alternative drilling platform. "We won't see the day when a conventional drill ship can go into these ice-covered areas," he points out, adding that he is getting plenty of interest from other scientists in the approach.   **A.W.**

N. DUBREUIL / LOOKATSCIENCES

The new face of the Arctic

Every summer the Arctic Ocean loses more ice — and it could all be gone within decades.
Quirin Schiermeier looks at how the vanishing summer ice affects those living in the north.

It was on Christmas day that Duane Smith first noticed that something weird was happening. When he and his family went to church, they did so in the rain.

That was in 1983. "We'd never ever seen anything like it," remembers Smith — then a little boy in Inuvik, a town of 3,500 people just north of the Arctic Circle, and now president for the Inuit Circumpolar Council (Canada), an indigenous people's organization. "Around Christmas it was supposed to be some 30 degrees below zero. None of our elders had any memory of such mild weather in winter."

It could just have been weird weather. In fact, it was a harbinger of things to come. Nowhere else on the planet is the current warming trend more pronounced than in the Arctic, and nowhere else does it seem to leave a deeper mark. The Arctic was a favoured site for early-warning systems during the cold war. Today, it is the early-warning system for climate change.

In recent years, researchers have started to pin down the details of what might happen to the Arctic as the planet warms. Although many of the specifics remain speculative (see 'How the Arctic might change', overleaf), everyone is certain that change is coming — and fast. "The Arctic is changing extremely abruptly on a geological time scale," says David Barber, a climatologist and sea-ice specialist at the University of Manitoba in Winnipeg, Canada. "There is no good historical analogue that could tell us what might happen."

Blame it on the sea ice. Unlike Antarctica — a continent surrounded by oceans — the Arctic is for the most part an ice-covered ocean, making it particularly vulnerable to climate change. Within several decades, the entire Arctic Ocean, including Hudson Bay and the countless channels between Canada's Arctic islands, could be free of ice during the summer months[1]. Palaeoclimatic evidence suggests that this has not been the case for at least the past 1 million years.

Less ice during the Arctic summer might not necessarily be all bad. New shipping channels and oil and gas regions could open up, for instance, and local hunters could get around by boat more easily (see 'Life in a warming world', overleaf). But the rate and magnitude of the changes are unprecedented, and the consequences are difficult to predict[2].

The amount of sea ice in the Arctic usually reaches its maximum — more than 14 million square kilometres in recent years — around the end of March. The slowly moving pack ice is separated from the immobile ice attached to the coastlines by the perennial 'circumpolar flaw lead' — a narrow corridor of open water that is rich in biological productivity and crucial for the heat exchange between the ocean and the atmosphere. This lead will be the focus of a multinational expedition led by Canada during this International Polar Year.

During the summer, ice melts and thins, reaching its minimum in September. The minimum extent of sea ice in the Arctic has decreased from a long-term average of more than 7 million square kilometres since 1979, to less than 6 million square kilometres in 2002 (ref. 3). Every year since, it has continued to drop or stay at near-record low levels. In
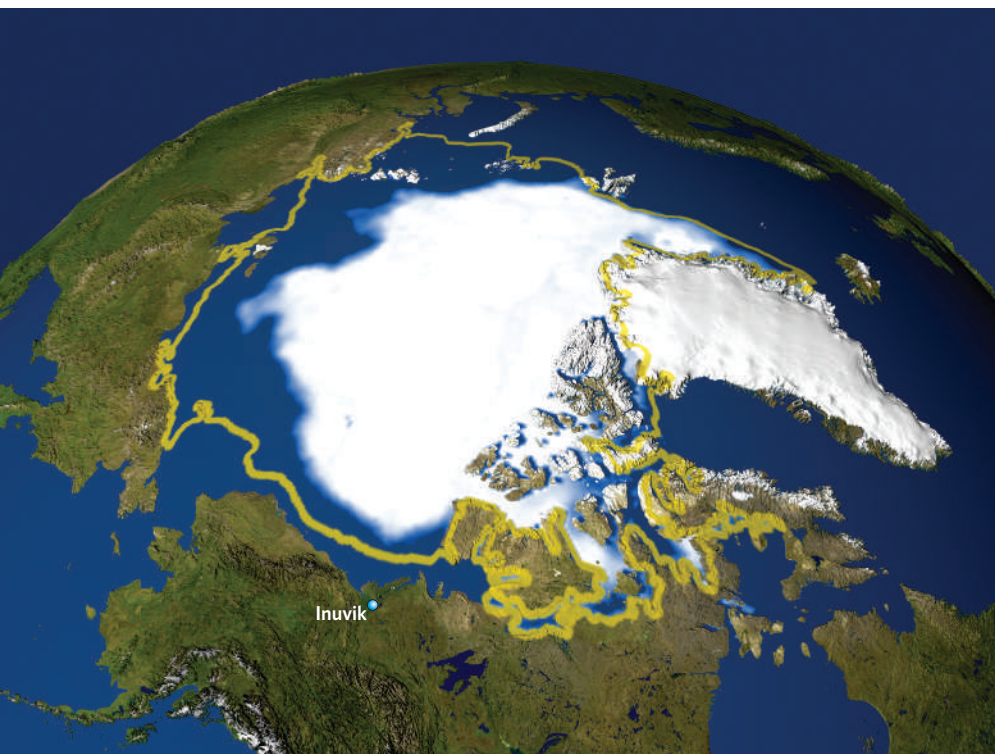
By 2005, summer ice coverage was only about three-quarters of the Artic's long-term average (outline).

September 2005, the Arctic was covered by just 5.32 million square kilometres of ice — the lowest yet.

As more and more ice disappears, a vicious cycle sets in. Ice reflects away a large fraction of incoming sunlight, whereas the darker open ocean absorbs more radiation. This 'albedo' effect is the main reason why the influence of warming is significantly more pronounced at high northern latitudes than in temperate or tropical regions. A reduced albedo in the Arctic affects the entire planet's energy balance, causing yet more energy to be absorbed in the darker waters.

A sizeable problem

The ice is not only shrinking in its area, but also in its depth, as recorded by submarines[4] and radar images from satellites. And it seems to be declining even in the winter[5]. If the ice continues to disappear at its current rate of nearly 9% per decade, the Arctic Ocean will be ice-free in September by 2060. But if, as some scientists suspect, the shrinking were to accelerate, this date will come forward 20 years to 2040.

This dire scenario is just one of seven computer simulations published in December by a team led by Marika Holland of the National Center for Atmospheric Research in Boulder, Colorado[1]. That particular simulation suggests that the summer Arctic sea ice will decrease

from 6 million square kilometres to 2 million square kilometres in the course of a decade. The ice that remains would be tucked along the coasts of Canada and Greenland, leaving the central Arctic basically free of ice by the end of the melting season, although the region would refreeze during the winter.

The receding ice cover could also affect large-scale patterns of ocean circulation. Ice, for instance, seems to be moving at an increasing

rate out of the Arctic Ocean through the Fram Strait to the east of Greenland and through the Canadian Arctic archipelago to Greenland's west. On average, the amount of ice moving out is around 10%, but during the winter of 2005 to 2006, a strong counterclockwise rotation pattern in the Arctic Ocean pushed about 40% of the pack ice into the warmer Atlantic waters. The events over the past two years, says Barber, are the first sign that the rates of ice export can change dramatically. If more storms start to enter the Arctic, as expected with the rising temperatures around the world, the pack ice will be broken up and potentially carried away more often.

Flow-on effects

Reduced sea-ice cover might also increase the influx of warm Pacific waters through the Bering Strait between Russia and Alaska. Koji Shimada, a physical oceanographer at the Institute of Observational Research for Global Change in Yokosuka, Japan, has suggested that because sea ice is starting to form later each year, the warmer Pacific waters will be able to flow into the Arctic[6]. Such ocean changes could dramatically affect the lives of Arctic organisms, from algae to fishes to mammals. "All indigenous animals are somehow tied to ice," says Jacqueline Grebmeier, a biological oceanographer at the University of Tennessee in Knoxville. "So the early-season retreat of ice will no doubt trigger cascading effects."

For instance, the thinning and retreat of sea ice allows more sunlight to penetrate the ocean earlier in the season, making surface waters richer and more biologically productive. Already, scientists have seen the dominant zooplankton in the Arctic — the small

HOW THE ARCTIC MIGHT CHANGE

5 years from now	10 years from now	20 years from now
The polar bear is listed as a threatened species.	The amount of multi-year pack ice decreases catastrophically.	The Arctic Ocean remains ice-free in September.
The UN Convention on the Law of the Sea adopts a protocol for the new polar ocean.	A new UN environmental regime for the Arctic comes into effect.	All known Russian offshore hydrocarbon deposits (oil, gas and gas condensate) are exploited commercially.
Development of hydrocarbon deposits starts in the eastern Kara Sea off Siberia.	Russia auctions offshore exploration licences for up to 20 billion tonnes of natural gas and petroleum.	Offshore oil and gas resources, unknown today, are being extensively developed.
The Galileo satellite system is used to combat illegal fisheries in Arctic waters.	Vector-borne diseases become more widespread among the Inuit as mosquitoes migrate to the Arctic.	Oil tankers and container ships sail the Northeast passage.
Norwegian oil companies commit themselves to being carbon neutral.	The Arctic cod is displaced by temperate fishes.	The United States and Russia undertake military manoeuvres in the polar sea.
All-season Canadian icebreakers patrol the eastern and western Arctic.	Canada installs high-frequency radar at the entrance to the Northwest passage to bolster its sovereignty.	

Subsistence farmers in Alaska adapt their practices to cope with the ever-changing climate.

Life in a warming world

Some 100,000 Inuit live in regions north of the Arctic Circle on land that their ancestors have hunted and fished for generations. They know, both from their own observations and from what scientists tell them, that their environment is a hot spot for global warming.

"We don't get much cold any more, spring is coming earlier, and ice conditions are getting unpredictable to the point of people falling through the ice and drowning," says Duane Smith, an Inuit leader from Inuvik, Canada. Yet many reject the notion that climate change is all bad.

"I feel that governments are panicking a bit," says Frank Pokiak, a native of Tuktoyaktuk, Alaska, who chairs the Inuvialuit Game Council. "People need to understand that we've been living with changes all our lives. Climate change is just another thing we need to adapt to. We may need to

harvest other species, perhaps grizzly bear, perhaps caribou, but we won't quit existing."

A warming Arctic has some advantages. The early breaking up of ice, for example, gives hunters a longer time to harvest beluga whales. Less ice would also facilitate hunting and travelling by boat.

On the other hand, regional hubs such as Inuvik depend on roads across the ice for food and other supplies. Hunting, transport and road safety will all be affected if the tundra turns into bogs earlier and freezes later. Increased dependence on helicopters has already notably increased the Inuits' cost of living, says Smith.

Many Inuit are working with scientists to help investigate the changes in the Arctic. Hunters and trappers, for instance, are being taught how to handle meteorological instruments. The Inuit have plenty to teach

the scientists as well; at one point, researchers were about to conclude that Arctic cod (*Arctogadus glacialis*) had all gone because stocks were nowhere to be found, but Inuit fishermen showed them where on Canada's Mackenzie shelf the fish had hidden.

But some things are more difficult for indigenous people to understand. Explaining to Inuit that they and their food could be contaminated with chemicals produced far in the south is challenging, says Louis Fortier from University Laval in Quebec. Their native language knows nothing of 'molecules' or 'chemistry'.

So, working with Inuit elders, linguists from ArcticNet, an interdisciplinary network of Canadian researchers, have produced a bilingual glossary for terms that relate to climate change. In it, for instance, the word for 'carbon' — a key term in the future of the Inuit — is illustrated as 'the soot of fire'. **Q.S.**

copepods *Calanus hyperboreus* and *C. glacialis* — replaced by their warmer-water, Atlantic cousins *C. finmarchicus*. Their move is thought to be related to the fact that the surface waters in the Arctic are now warmer and so more algal food can be found there.

Mix and match

This 'Atlantification' of the Arctic comes at the expense of species composition and biodiversity, says Louis Fortier, a biological oceanographer at the University Laval in Quebec. As warmer water enters the Arctic basin, generalist species from temperature latitudes may move in and out-compete the Arctic specialists. In Canada's Hudson Bay, for instance, black guillemots (*Cepphus grylle*) used to feed exclusively on Arctic cod (*Arctogadus glacialis*). In recent years, scientists have found that up to half of the birds' stomach contents consisted of small capelin fish (*Mallotus villosus*), which are common around Iceland. Meanwhile, Arctic cod are increasingly being replaced by the larger Atlantic cod (*Gadus morhua*) as the dominant species in some parts of the Arctic Ocean.

Seals, whales and polar bears all feed on the smaller, easier-to-catch Arctic cod. The shrinking of the sea-ice habitat and the shifting ecosystem have already affected the behaviour of the most emblematic Arctic species — polar bears, walruses and ringed seals. The Pacific walrus (*Odobensus rosmanus divergens*), for instance, is being forced to feed in deeper waters as the sea ice retreats off the coasts of continents, and mother walruses seem to be separated from their calves more often as sea ice continues to fragment[7]. And polar bears are being considered for listing in the United States as a threatened species because of their disappearing sea-ice habitat.

Humans, too, are preparing for the change. Subsistence hunters are learning to follow mammal and fish populations into new areas as the ecosystem shifts[8]. And Inuit leaders such as Smith are realizing how they will have to adapt — and are even looking forward to it. "Our people have understood ice conditions for ages," he says. "Now we're keen to learn how global warming will change our world." ■
Quirin Schiermeier is a correspondent in *Nature*'s Munich office.

1. Holland, M. M., Bitz, C. M. & Tremblay, B., *Geophys. Res. Lett.* **33,** L23503 (2006).
2. Arctic Climate Impact Assessment *Impacts of a Warming Arctic* (Cambridge Univ. Press, 2004).
3. Stroeve, J. C. *et al. Geophys. Res. Lett.* **32,** L04501 (2005).
4. Rothrock, D. A. *et al. J. Geophys. Res.* **108,** 3083 (2003).
5. Comiso, J. C. *Geophys. Res. Lett.* **33,** L18504 (2006).
6. Shimada, K. *et al. Geophys. Res. Lett.* **33,** L08605 (2006).
7. Cooper, L. M. *et al. Aquat. Mamm.* **32,** 98–102 (2006).
8. Grebmeier, J. W. *et al. Science* **311,** 1461–1464 (2006).

Blakemore steps down from Medical Research Council

Neuroscientist Colin Blakemore is to give up his position as head of Britain's Medical Research Council (MRC) when his present contract expires this September.

Blakemore, who has been in post since 2003, told *Nature* that he has decided not to apply for a second term because the MRC will suffer a temporary loss of independence while new advisory boards are established to oversee UK health funding. The MRC distributes around £500 million (US$960 million) of government money per year.

He departs with support from many of the researchers he funds. Over-optimistic spending had left the council short of money before Blakemore's arrival; he stabilized the funding situation and simplified grant procedures.

Sacked professor launches legal action to regain job

A professor dismissed by the University of Tokyo in Japan said on 2 March that he will take the university to court to demand his job back.

Kazunari Taira was fired in December after an investigation found "no reproducibility and no credibility" in four RNA papers from his lab. Also sacked was the researcher who had been in charge of the questioned experiments (see *Nature* 445, 12; 2007). Both deny involvement in misconduct.

The university criticized Taira for "neglect on appropriate supervision and lab management". But Taira argues that the case could put researchers off running labs and wants the university to restore his professorship and resume paying his salary. A spokesperson for the university says that it is considering its response to the lawsuit. Hearings for a provisional ruling will begin this month.

Nobel medal stolen from cabinet finds its way home

A Nobel medal swiped from the University of California, Berkeley, was returned to officials in a brief ceremony on 7 March.

The solid gold medallion, awarded to physicist Ernest Lawrence in 1939 for the invention of the cyclotron, was stolen from its display case at the end of February. It has an estimated monetary value of $4,200.

Scientists plan defence against asteroids

There's no doubt that a space rock slamming into Earth could cause substantial damage, but exactly what humans should do about the threat has not yet been decided. That's why scientists gathered for the Planetary Defense Conference in Washington DC on 5–8 March. Their aim was to compose a white paper on the subject — the first to be mandated by the US Congress.

Scientists at the conference said that it would cost about US$1 billion to find at least 90% of the 20,000 estimated potential Earth-killers by 2020, and discussed how a space rock on a collision course might be deflected. Options range from using spacecraft as 'tugboats' to drag an object into a new orbit, to proposals that rely on nuclear detonations to knock a rock off target — similar to the strategy used against a comet in the 1998 film *Deep Impact* (pictured). The white paper will be published at www.aero.org/conferences/planetarydefense.

Authorities last week charged a biology student at the Berkeley campus with the theft. The 22-year-old worked at the university's Lawrence Hall of Science, where the medal was stored. He told police that he used a key to open the case, taking the 200-gram medal on a whim. Plans are now under way to house the medal in a more secure manner.

Wildlife agency accused of gagging scientists

The US Fish and Wildlife Service is fending off accusations of scientific censorship over a memo that outlines restrictions on who can say what about climate change, sea ice and polar bears.

The memo states that agency personnel travelling in northern countries should indicate a spokesperson and provide assurances that they "understand the Administration's position on these issues". A travel request for a non-spokesperson, included in the memo as an example, said that the individual "will not be speaking on or responding to these issues".

Hugh Vickery, a spokesman at the Department of the Interior, the service's parent agency, says the memo was "badly worded" and really only applied to discussions of policy. "It's a reminder not to get into areas where they shouldn't go." The polar bear has become political because is it being considered for a slot on the US endangered species list, thanks to the threat of melting sea ice.

The diaries of Darwin's wife debut online

For a unique glimpse into the family life of Charles Darwin, historians can now delve online into the diaries of his wife Emma.

Emma noted in pocket diaries details of her everyday life such as shopping costs, dinner parties and the illnesses of her children and husband. Spanning from 1824 — before the pair's marriage in 1839 — to 1896, the diaries belong to Darwin's great grandson and are looked after by the Darwin Archive at the University of Cambridge, UK. Images of the diaries' pages were made available to browse at http://darwin-online.org.uk/EmmaDiaries.html on 6 March.

The diaries record the visits of friends and scientists to the couple's house. For example, an entry in April 1856 says "Lyells came". A note in Darwin's own journal, dated a few weeks later, tells what happened after the visit by Charles Lyell and his wife: "Began by Lyells advice writing species sketch."

Correction
The News Feature 'The new face of the Arctic' (*Nature* **446,** 133–135; 2007) mistakenly located the community of Tuktoyaktut in Alaska. It is in Canada's Northwest Territories.

Abderhalden's fraud still wins him some supporters

SIR — Your Misconduct Special (*Nature* **445**, 240–245; 2007) calls South Korean stem-cell biologist Woo Suk Hwang "arguably the highest-profile fraudster ever". A look into the history of fraud in the biomedical sciences reveals at least one other strong candidate for this title.

In the area of human reproductive biology, the Korean scandal is overshadowed by the case of the influential German physiologist Emil Abderhalden (1877–1950) and the non-existent *Abwehrfermente* or 'defence enzymes' he claimed to have discovered. The shocking story of his fraudulent work, over a long period, has been told by Ute Deichmann and Benno Müller-Hill (*Nature* **393**, 109–111; 1998).

Briefly, Abderhalden — a powerful and influential scientist — published a first paper on his "most important discovery" in 1909, and a widely read and translated book on the subject followed in 1912 (E. Abderhalden *Abwehrfermente: Die Abderhaldensche Reaktion* Theodor Steinkopff, 7th edn, 1944). The 'Abderhalden reaction' was used as a pregnancy test, and to treat various diseases. From 1914 on, biochemists tried to repeat Abderhalden's experiments, but failed to achieve this. A number of experts published their negative findings, but Abderhalden continued to publish fabricated data until his death.

In 1998, Deichmann and Müller-Hill asked whether such a fraud could happen again. Hwang's case showed that it could — but that was, at least, quickly exposed and denounced.

Yet Abderhalden is still regarded by some as one of the founding fathers of modern biochemistry (see M. Kaasch *Acta Hist. Leopold.* **36**, 145–210; 2000). The nomination database for the Nobel Prize in Physiology or Medicine reveals that, between 1913 and 1936, he was listed 49 times as a nominee for this most prestigious prize. Sixteen times, the motivation was Abderhalden's "work on defence enzymes". Twice, his pregnancy test is highlighted, and in one case "all the work Abderhalden has done" is listed in this database (see http://nobelprize.org/nomination/medicine/database.html). There is no mention of his fraud on the Nobel site. These numerous nominations, combined with Abderhalden's despotism, narcissism and influence as a science manager, prevented his work from being rejected as an outright scientific fraud.

Worst of all, Abderhalden's myth is still alive. For instance, on the German site of the Internet encyclopaedia Wikipedia, Emil Abderhalden was until this year listed as an "important physiologist and discoverer of the specific Abderhaldensche Abwehrfermente — he developed the first pregnancy test". Similar admiring descriptions can still be found in the latest editions of German lexica such as *Die Zeit: Das Lexikon 2005* (www.zeit-lexikon.de/index.php). The authors refer to Abderhalden's book, which remains available in libraries and second-hand, and is still taken seriously in the popular literature on biomedicine.

U. Kutschera
Institute of Biology, University of Kassel,
Heinrich-Plett-Strasse 40,
D-34109 Kassel, Germany

A Moon base offers more than just a good view

SIR — Your Editorial "Brave blue world" (*Nature* **445**, 459; 2007) correctly points out that Antarctic science "has been central to the great project of understanding the changes that humans are inflicting upon the Earth". You go on to say that "the only input that lunar activity will provide for the study of Earth is the iconic and inspiring sight of a blue planet in a black sky over a grey desert".

But when Antarctic exploration began, the concept of global warming that had been introduced by G. S. Callendar[1] in 1938 was only beginning to be taken seriously[2]. Studies in the 1950s focused mainly on ocean turnover of carbon dioxide. That an Antarctic base would reveal humanity's contribution to climate change with such clarity was not a driver for building a base, and ultimately many findings were serendipitous — we certainly were not looking for an ozone hole, for example.

If the only goal of an Antarctic base was to gather atmospheric data, it would have been easier and much cheaper to use aircraft rather than build a base. Similarly, if the only goal of a Moon base is to observe Earth, it is cheaper and easier to use satellites. This misses the main point of establishing a distant outpost. In general, if one picks a single function of the outpost, it is possible to find a cheaper and easier way to perform that particular function. But the multiple goals of a broad scientific expedition help each other succeed in many ways. Antarctic discoveries were made because we established long-term monitoring in a scientifically interesting place with no guarantees of the ultimate importance of the science.

Similarly, the Moon is a scientifically interesting place to explore and, although there are no guarantees, long-term monitoring of the entire Earth, as visible from the Moon, is a pretty good bet to provide important discoveries that we are unable to anticipate at present. Just as the synergies of exploration, oceanic, atmospheric, geological, biological and meteoritic investigations combined to yield rich results from Antarctica, we will one day be amazed by the wealth of knowledge gained from the Moon.

Dan Barry
NASA astronaut (retired), Denbar Robotics,
South Hadley, Massachusetts 01075, USA.

1. Callendar, G. S. *Q. J. R. Meteorol. Soc.* **64**, 223–240 (1938).
2. Revelle, R. & Seuss, H. E. *Tellus* **9**, 18–27 (1957).

Fossils: new journal will oppose illegal trade

SIR — The Association of Applied Paleontological Sciences (AAPS), of which I am president, changed its name in 2002 from the American Association of Paleontological Suppliers, to reflect our international and broad-based membership. It had nothing to do with the impending launch of the *Journal of Paleontological Sciences* (*JPS*), as claimed in your News story "Palaeontology journal will 'fuel black market'" (*Nature* **445**, 234–235; 2007, and Correspondence *Nature* **445**, 709; 2007).

The AAPS actively supports efforts to locate and retrieve stolen specimens. The journal's submission guidelines state that the *JPS* will not publish fossils that cannot be legally exported from their home country, or where ownership cannot be verified. The view attributed to Mark Goodwin of the University of California Museum of Paleontology in Berkeley — that the new journal will encourage commercial fossil collection and hence increased illegal collecting activity and trade on public lands — is incorrect. Authors who publish in the *JPS* must abide by the AAPS code of ethics, and must agree to make their published specimen(s) available for study, as do many other scientific publications (see www.aaps-journal.org).

Commercial fossil trading in the United States started with quarrymen in New Jersey selling fossils to Joseph Leidy during the mid-1800s. By the 1870s, professional collectors were busy filling museums with dinosaurs and other fossils, by accepting the risks of exploration, discovery and excavation, then selling their discoveries, and in some cases collecting fossils on a contract basis. Visit any number of prestigious institutions and you will see magnificent displays whose very existence is owed to professional collectors.

The tax deduction allowed for fossil donations under US tax law (which is not a government policy) is not "a bid to confront these activities", but is a long-standing policy having nothing to do with fossils *per se*. You can donate your Grandpa's spittoon to a museum and get the same write-off. The donation doesn't have to be to a museum: any 501(C-3) non-profit corporation can accept tax-deductible donations.

Michael Triebold
Triebold Paleontology and Rocky Mountain Dinosaur Resource Center, 201 South Fairview Street, Woodland Park, Colorado 80863, USA

COMMENTARY

Keeping faith with trial volunteers

How best to serve patients' interests in large clinical trials? **Martine Piccart**, **Aron Goldhirsch** and their colleagues argue that maintaining academic independence is essential to early breast cancer trials.

How do patients who sign up for clinical trials know that the right questions are being asked and that the data support the reported answers to these questions? For many patient groups, transparency in study design, data collection and analysis, and full publication of results are issues of paramount importance. Concerns have been raised in recent years about the misreporting of trial results and we note a push by pharmaceutical companies for greater control of clinical trials and data, outside the framework of academia. In our view, a more equal partnership between academic researchers and the pharmaceutical industry is better for patients, especially those in early stages of disease for whom over-treatment and adverse side effects are important considerations.

The substantial progress in treatment of early breast cancer, reflected in falling mortality rates in many countries, is the direct result of more than half a century of sustained commitment to randomized clinical trials[1]. The implicit ethical contract between the investigators and the many patients who participate in trials has made this progress possible. The prime objective of clinical trials is to improve the outcomes for patients with the disease, and no consideration of commercial or academic advantage should cloud the collection or interpretation of data from such trials.

Conflicts of interest

Modern clinical trials often seek subtle but important improvements in outcomes, for example with adjuvant therapies the intended outcome is to keep cancer from recurring. Trials are also necessary to define the frequency and severity of any adverse reactions to treatment, which affect the patient's quality of life and influence future treatment decisions. Moreover, trials increasingly tailor treatment questions to specific subpopulations as investigators learn more about tumour biology or patient drug tolerance. For example, one trial may be designed to test adjuvant chemotherapy in patients older than 70, and another the adequacy of endocrine treatment in a specific biologically defined target group[2,3]. Assessing the effectiveness of such therapies requires trials, usually multicentered and multinational, with very large numbers of participants (several thousands or more).

> **"Trials designed and controlled purely by drug companies may fail the best interests of patients."**

Who should collect, control and analyse clinical-trial data: academic investigators or drug companies?

Trials of adjuvant therapies in early breast cancer involve treating patients whose disease has not relapsed following local treatment with surgery, or in some cases radiotherapy, and therefore individuals have no evident disease that can be assessed. Instead, group comparisons after an adequate period of follow-up — usually in the range of 2 to 5 years — are needed to assess average outcomes, and to explore patient or tumour differences that may influence treatment benefit[4]. Typical agents used in adjuvant treatments include hormonal therapies such as tamoxifen, or chemotherapy alongside newer targeted agents such as trastuzumab. Without the clinical trial process it is impossible to determine the benefits of experimental treatments for individual patients. Trial analyses will ultimately feed into individual patients' informed decisions on future treatments.

Such large modern trials need financial resources that currently exceed those available to academic researchers from governments and philanthropy alone. Partnership with pharmaceutical industry, often for the supply of the drug, is usually essential. But how should such partnerships work? Contrary to the interests of academics, and we believe to those of patients, pharmaceutical companies are increasingly attempting to recruit academic investigators to conduct adjuvant trials in which the data will be controlled by the company outside the framework of a research cooperative group or a network of academic centres[5].

It is important to understand the conflicts of interest inherent in the clinical trials process. At times the interests of the pharmaceutical industry and those of academic investigators overlap, and at other times they diverge, although both are legitimate. For example, the pharmaceutical partner might wish to establish or extend regulatory approval for a commercial version of the tested product; whereas the academic researchers might be interested in the publication of the trial results, which help to advance their careers.

In some cases, academics will wish to explore trial results in detail to derive hypotheses that may eventually lead to even more precisely tailored treatments for future patients. Such efforts often require translational research investment and prolonged follow up, beyond that needed for commercial drug registration. So, views on funding levels and trial length may differ between the partners[6]. However, all parties are interested in the success of the experiment.

A partnership with the pharmaceutical industry is essential for the supply of drugs to be tested in clinical trials.

No one should forget that commercial success of trials is necessary to ongoing pharmaceutical research investment. But we believe trials designed and controlled purely by pharmaceutical companies, particularly relating to new and highly expensive pharmaceutical agents, may fail the best interests of patients in several ways. We call upon academic investigators recruited for such trials to consider the following issues carefully before they agree to participate.

Data control

First, if a trial is focused on answering a purely commercial question, vital opportunities to answer other important questions related to the care of patients and to biological understanding may be lost. Second, trial design can be distorted by commercial interest, for example, requiring an arbitrary duration of treatment, rather than focusing on the optimal treatment duration for patient benefit. We note an increasing tendency, especially in pharmaceutically controlled trials, to withdraw funding or cease follow-up studies after commercial endpoints have been satisfied.

Regrettably, cases have been recorded in which data held by the pharmaceutical industry have allegedly been withheld for commercial reasons[7]. Data control entirely within a commercial organization may enhance the temptation to delay or suppress unwelcome findings. For example, large trials designed to define a subset of the patient population that benefit most from a treatment can run counter to the interests of a drug company wishing to maximize the number of potential patients for a new treatment. In such cases, control of data by the drug company would not be in the best interests of patients. When such activities

become public they affect the credibility of all clinical trials as well as that of the commercial organizations involved[8]. In the long run, concerns about the control and reporting of data may deter eligible patients from participating in future trials.

Indeed, by seeking to control clinical trial data, the pharmaceutical industry is ignoring moves towards greater data transparency, as increasingly demanded by institutional review boards (IRBs) and informed consumers. We support calls by the World Medical Association[9] and others to expand disclosure of funding sources and financial conflicts of interest to potential trial participants. We also welcome the possibility that in the near future consumers and IRBs may insist on full disclosure of who collects, controls and analyses trial data. It is essential that the data to explore and report all possible findings from clinical trials are transparent and available to the academic investigators and thus ultimately to consumers.

> "Concerns about the control and reporting of data may deter eligible patients from participating in trials."

Win–win situation

We believe that undesirable outcomes can be avoided by pursuing a better partnership between academia and industry. We note several recent examples of such successful partnerships in large-scale breast cancer adjuvant trials, including MA.17, HERA, BIG 1-98 and IES[10–13]. In these trials the clinical trial database was held by academic researchers who limited industry's access until the trial outcomes were reached. All relevant parties had access to safety data (adverse events and so on), but outcome data were analysed by an independent statistician, and while the trial was ongoing, shown only to an independent data monitoring committee (IDMC), which had the task of ensuring patient safety and the timely release of positive

or negative outcomes. The IDMC members for these trials were independent from industry, usually involving academic researchers knowledgeable about clinical trials, but not involved in the specific trials concerned, plus a consumer representative and at least one statistician.

In each case, once the primary endpoints of the trial have been reached, the database is transferred to the pharmaceutical company for commercial registration purposes. This model reduces commercial bias and conflicts of interest between the parties involved, while ensuring the protection of patients. In our view, it yields a win–win situation resulting in commercial registration of products, academic publications, and last but not least, hopefully better outcomes for patient treatments.

We firmly believe that such a model provides the best basis of fruitful current and future collaboration between academia and industry. Simultaneously, it maintains sacrosanct the overriding interests of the participating patients and those who will follow them. Only by ensuring untrammelled access to long-term information, both the good and the bad, can we conduct clinical trials in a credible manner. This access will ensure that those patients who consent to participate in them maintain faith in the clinical trial process. ■

Martine Piccart and Aron Goldhirsch are at the Breast International Group. Their co-authors are William Wood of The Breast Cancer Intergroup, Kathleen Pritchard of The Early Breast Cancer Trialists Collaborative Group, José Baselga of Grupo Espanol de Estudio y Tratamiento Tumores Solidos, Linda Reaby of The Australian New Zealand Breast Cancer Trials Group Consumer Advisory Panel, Ingrid Kössler and Stella Kyriakides of The European Breast Cancer Coalition, Larry Norton of The Breast Cancer Research Foundation and Alan Coates of The International Breast Cancer Study Group.

1. Early Breast Cancer Trialists' Collaborative Group *Lancet* **365,** 1687–717 (2005).
2. Giordano, S. H., Duan, Z., Kuo, Y. F., Hortobagyi, G. N. & Goodwin, J. S. *J. Clin. Oncol.* **18,** 2750–2756 (2006).
3. Regan, M. M. & Gelber, R. D. *Breast* **14,** 582–593 (2005).
4. NIH *Adjuvant Therapy for Breast Cancer* available at http://consensus.nih.gov/2000/2000AdjuvantTherapyBreast Cancer114html.htm (2000).
5. Mello, M. M., Clarridge, B. R. & Studdert, D. M. *N. Engl. J. Med.* **352,** 2202–2210 (2005).
6. Gilpin, K. N. *New York Times* (2 June 2004).
7. Angell, M. *The New York Review of Books* **53**(10), (2006).
8. Drazen, J. M. *N. Engl. J. Med.* **347,** 1362–1363 (2002).
9. World Medical Association *Declaration of Helsinki* (www.wma.net/e/policy/b3.htm).
10. Goss, P. E. *et al. N. Engl. J. Med.* **349,** 1793–1802 (2003).
11. Piccart-Gebhart, M. J. *et al. N. Engl. J. Med.* **353,** 1659–1672 (2005).
12. Thürlimann, B. *et al. N. Engl. J. Med.* **353,** 2747–2757 (2005).
13. Coombes, R. C. *et al. N. Engl. J. Med.* **350,** 1081–1092 (2004).

BOOKS & ARTS

Immortality of a kind

The ability to grow human cells in the laboratory created paradoxes of personal identity.

Culturing Life: How Cells Became Technologies
by Hannah Landecker
Harvard University Press: 2007. 276 pp.
$35, £22.95, €32.30

Nick Hopwood

In the flood of instant comment on cloning and stem cells, we need the longer and deeper views of cellular technologies that only history can provide. Historians of science have written much about the nineteenth-century advent of cell theory, but genes and molecules stole the limelight in the twentieth. We have a first-hand account of the history of somatic-cell genetics (*The Cells of the Body* by Henry Harris; Cold Spring Harbor Laboratory Press, 1995), a rich study of the adoption of the electron microscope (Nicolas Rasmussen's *Picture Control*; Stanford University Press, 1997), and a philosophically driven interpretation of the rise of cell biology (*Discovering Cell Mechanisms* by William Bechtel; Cambridge University Press, 2006). But there has been no extended history of tissue culture — the technique, which underpins most biomedicine today, for growing vertebrate cells in the laboratory as if they were independent microorganisms. With five chapters tackling key episodes up to 1970, *Culturing Life* by Hannah Landecker is a small book that does much to fill that large gap.

Landecker adopts a powerful approach from recent science studies: she takes routine practices of observation and manipulation very seriously indeed. This might sound dull, and not everyone would choose to spend years poring over methods sections and manuals. What converts base method into golden insight is the anthropologist's eye for the strangeness, and thus the historical significance, of techniques that practitioners soon took for granted. Landecker identifies fascinating novelties in the autonomy, plasticity and time relations of cultured cells. She shows how, long before Dolly was born, such mundane technologies as flasks, tubes, nutrient media, freezers and culture collections created radically new and challenging forms of life.

Tissue culture was pioneered in the early twentieth century by scientists frustrated with 'fix, slice and stain' histology and its obligatory detour via the cadaver. To solve difficult problems — the process of nerve outgrowth and the origin of the heartbeat — they learned from bacteriology how to culture living cells outside

the body and so see them more directly. Observation was still highly mediated. Landecker reveals how time-lapse microcinematography made once-static entities move and change.

The drive to manipulate cells *in vitro* was about distinguishing inherent limits from technical obstacles that could be overcome. Yes, cells could divide, it was soon shown, but for how long? Between the world wars, the French-American surgeon Alexis Carrel sensationally claimed immortality. Wide audiences were told that, with enough food, his culture of chick embryo cells would grow larger than the Sun. He was believed, Landecker suggests, because his claim fit with a prevalent ideal of biological engineering. It would be interesting to go further and explore how, in the era of testicular transplants to restore the failing powers of rich old men, cellular immortality was bound up with the whole-organism biology of death, ageing and rejuvenation.

Carrel was plausible because experiments were restricted to a few laboratories with their own distinctive cultures, in every sense. After the Second World War, the campaign to mass-produce polio vaccine led to tissue culture being practised on a far larger scale and applied

FIREFLY PRODUCTIONS/CORBIS

Life in the lab: the ability to store and culture human cells led to the creation of the HeLa cell line.

to the previously recalcitrant human cells. Techniques and reagents were standardized, and so, like other model organisms, were the cells. Freezing and clonal cultures promoted the distribution of established lines and liberated cells and researchers from the constraints of space and time. Life could now be started, stopped, stored, split and its different stages juxtaposed. With the finding that normal somatic cells can divide only a fixed number of times, Carrel's claim of cellular immortality was rejected, but some cells and some people still achieved immortality of a kind.

Landecker interprets the various stories about Henrietta Lacks — a black American who died of cervical cancer not knowing that her biopsy had been turned into the permanent HeLa cell line — as attempts to negotiate the paradoxes of personal identity in the biomedical age. Optimism in the 1950s about having a laboratory afterlife of service to science gave way in the late 1960s and 1970s to racially charged fears of contamination with these by then ubiquitous cells; now, bioethical tales of overdue recognition are dominant. Landecker brings out the Lacks stories' obsession with 'what she would weigh today' — the unsettling

phenomenon that more living, reproducing matter has been generated from a body than it ever contained.

Lay readers will appreciate the effort that kept Landecker's scholarly and original book short and accessible. There is inevitably more for specialist historians to do. The epilogue, which interprets the cloning of adult mammals as dependent on freezing and synchronizing cells, whets the appetite for a fuller discussion of cell-cycle work. It also raises the larger question of how the histories of somatic cells and of gametes have intertwined. Answering it, and more generally gaining a sense of the place and status of cell culture in biology, would have taken the book beyond journal and newspaper articles to the realm of textbooks and other synthetic works. But these are suggestions for research that happily can now build on Landecker's stimulating reconstruction of the cultures that gave us cultured cells. ■

Nick Hopwood is in the Department of History and Philosophy of Science, University of Cambridge, Cambridge CB2 3RH, UK, and is co-editor of Models: The Third Dimension of Science (Stanford University Press, 2004).

A scientific symphony

Harmonious Triads: Physicists, Musicians, and Instrument Makers in Nineteenth-Century Germany
by Myles W. Jackson
MIT Press: 2006. 368 pp. $40, £25.95

Peter Pesic

Why are so many scientists musical and so many musicians scientific? This relationship goes back to antiquity, but remains largely unexplored. In his book Harmonious Triads, Myles Jackson, a historian of science and an accomplished cellist, examines nineteenth-century Germany, where science and music interacted with particular intensity.

He begins with the physicist Ernst Chladni, who in the 1780s sprinkled sand on vibrating plates to produce fascinating images of wave motion. Jackson relates this exemplary demonstration to Chladni's novel musical instruments (have you heard of his euphone or clavicylinder?). These instruments were informed by the physics of vibrating bodies, which had both musical and commercial possibilities. But musical observations also informed science: for example, Chladni's production of longitudinal vibrations led to the intensive study of their properties, which was crucial for the development of wave theory.

Chladni considered his demonstrations not entertainment but Bildung, the quest for education and personal edification so important in German thought. Likewise, Jackson depicts German scientists and physicians using choral singing as a central unifying activity at their meetings, celebrating their camaraderie while augmenting their musicality and fuelling their patriotism. Alexander von Humboldt once invited the composer Felix Mendelssohn to write a festive cantata, which you can hear on the book's website (http://mitpress.mit.edu). Why do our scientific meetings no longer include 'singing savants'?

Music celebrated science, but the conflict between organic and mechanical views lay behind Goethe's distrust of newtonian science, as well as the musical automata animating E. T. A. Hoffmann's unforgettable stories.

Nothing less than the soul was at stake: Jackson emphasizes that "audiences did not want to be entertained; they now wished to be moved". Accordingly, builders sought to make pipe organs more expressive by enabling them to swell in volume, despite the difficulty of making the sound louder without its pitch rising. This problem led physicist Wilhelm Weber and others to important research on the speed of sound and the specific heat of various gases. By then, however, the taste for expressively swelling reed organs had (mercifully) diminished.

In the process, both musical and physical developments had led to an ever-increasing emphasis on precision and standardization. Jackson's description of the struggle over an international pitch for concerts discloses a whole comédie humaine. Who would tune the concert of Europe? Each nation vied for pre-eminence by insisting on its own pitch
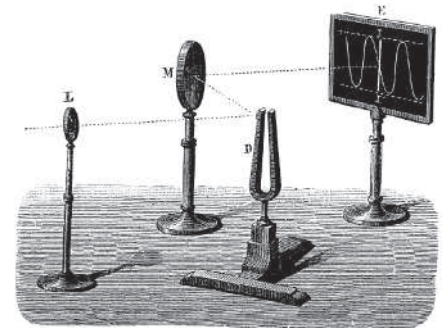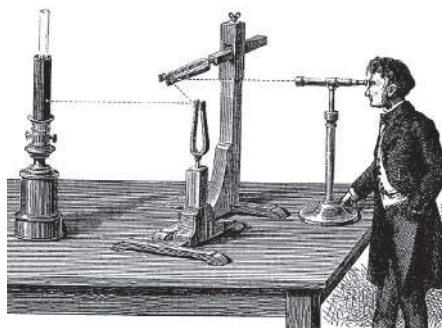
Sound idea: Jules Lissajous used light and mirrors to increase the precision of tuning forks.

standard, from Paris (where A above middle C was tuned to 435 Hz) to London (A455); by comparison, Mozart's own tuning fork sounded A422. An international conference convened in Vienna in 1885 chose the Parisian pitch, using arguments guided more by diplomatic finesse than musical purity. As one contemporary musician put it, using a higher pitch standard destroyed the "effect and character of ancient music — of the masterpieces of Mozart, Gluck, and Beethoven", who expected a lower pitch standard and did not wish their singers to strain a semitone past the pitches they had intended. Recent 'authentic' performances at the older, lower pitch standard have tried to reverse this trend.

Small comfort, then, that these new tuning forks were regulated by the research of physicists Jules Lissajous and Josef Stefan. Nor was the imposition of equal temperament (artificially equalizing the size of all semitones) an unmitigated boon, despite its simplicity and advantages in scientific eyes. Jackson's account of "the fetish of precision" in temperament describes how Johann Heinrich Scheibler and others produced increasingly accurate forks that enabled the tuning of keyboard instruments with unprecedented precision. This erased the earlier unequal temperaments used by J. S. Bach and his successors, in which each key had an individual character.

No less controversial was the development of the metronome. At first Antonio Salieri hailed the machine as "the true interpreter of the ideas and feelings of every composer". But soon its mechanical rigidity seemed only "a dumb thing; one must feel the tempi," as Beethoven put it.

At least the metronome could drill novices into developing a steady beat. Jackson concludes with the parallel development of other pedagogic torture devices that held a piano student's hand in the correct position using guide rails, such as Johann Bernhard Logier's chiroplast. These increasingly popular mechanical aids accompanied the spread of piano-playing as an indispensible bourgeois accomplishment, along with an idea of virtuosity that encompassed sheer speed, rather than depth of expression. Jackson concludes with fin de siècle controversies over how a pianist touching a key can create ineffable results beyond the scope of a mere mechanism.

Jackson brings forward both harmony and tension between science and music, for "the freedom of the individual to cultivate his or her own character and taste, the role of the State in defining those attributes, and the relationship between the organic and the mechanical were at stake." MIT Press should be commended for producing this beautiful volume. Jackson's outstanding book is an essential source for everyone interested in the relationship between music, technology and science. ■

Peter Pesic is tutor and musician-in-residence at St John's College, Santa Fe, New Mexico 87505, USA.

Life at the poles

Despite its harsh climate, the Arctic Circle is home to a rich fauna, including the male ptarmigan pictured here. A year-round resident of the Arctic, this bird retains his summer plumage on his head, with white winter feathers on his body. In *A Complete Guide to Arctic Wildlife* (A&C Black, £40), Richard Sale provides an extensive field guide to the birds and mammals that inhabit this harsh and unforgiving environment. The book also provides an introduction to the geology and geography of the region, shows how this has shaped the organisms that live there, and discusses the impact of humans on this fragile environment.

The World of the Polar Bear (Firefly, £29.99, $45) features a collection of stunning photographs by Norbert Rosing. The book explores a year in the life of the polar bear and the environment it inhabits.

Surviving Antarctica by David N. Thomas, to be published by Natural History Museum Publishing next month (£9.99), examines how endemic inhabitants and human visitors at the other pole survive in one of the most inhospitable regions on the planet.

P. MICHELSON

The changing view of autism

Unstrange Minds: Remapping the World of Autism
by Roy Richard Grinker
Basic Books: 2007. 304 pp. $26

Lorna Wing

The term 'autism' was coined by Leo Kanner in 1943 as a label for a disordered pattern of behaviour present from infancy that he and his colleague Leon Eisenberg had observed. They defined it very narrowly, as a profound lack of affective contact with other people, combined with an obsessive desire for the preservation of 'sameness' that results in elaborate, repetitive routines.

In *Unstrange Minds*, Roy Grinker combines his experiences as an anthropologist and as a father of a teenage daughter with autism, Isabel, to provide a fresh view of the challenges posed by this condition. He interweaves three themes: first, the evolution of ideas about the nature of autism and the effect of these changes on studies of prevalence; second, the intimate connection between cultural attitudes to autism and the quality of life for those affected and their families; and third, the story of his own daughter's development and the struggles that he and his wife have experienced in the United States to ensure that Isabel has the support and education she needs.

Grinker's discussion of the changing concepts of autism and its prevalence is highly topical. Some people believe that the number of children with autism has increased at an alarming rate since the 1990s and describe this as an 'autism epidemic'. But many others think that most, if not all, of the rise in prevalence reported in epidemiological studies can be explained by increasing awareness and a widening of diagnostic criteria. Feelings run high, especially among those who believe that, despite the lack of convincing scientific evidence, the increase is due to environmental causes such as pollution or a combined vaccination against mumps, measles and rubella (the MMR vaccine).

Grinker firmly rejects this view and rebuts any suggestion of an autism epidemic. He gives a compelling account of the complex way in which our understanding of the nature of autistic conditions has evolved. Children and adults with autistic behaviour have always existed, but in the past they were considered to be suffering from other conditions such as mental retardation, behaviour disorder or, in some cultures, some sort of malign influence. Once autism was identified as a childhood condition, researchers, clinicians and parents all contributed to the changes in diagnostic criteria.

Autism is no longer seen simply as Kanner's precisely defined, very rare condition, affecting as few as one or two children in every 10,000 and separate from all other childhood disorders. It is viewed instead as part of a spectrum of disorders of brain development causing impairments in social interaction, social communication and social imagination, along with a limited, repetitive pattern of behaviour. This spectrum has very wide manifestations that overlap with other developmental disorders and occur in people of any level of ability. The best evidence suggests a prevalence of around 10 in 1,000, even higher than the 6 in 1,000 quoted by the author. Grinker sees the higher reported prevalence as good news, as it indicates growing understanding and acceptance.

Grinker's exploration of cultural differences in attitudes to autism is very moving. He describes the experiences of individual children and their families in South Africa (a Zulu family), the United States, India and South Korea. Each culture has very different explanations for the strange behaviour of these children. However, in all societies, the main burden of caring for a child with autism is borne by the mother. During much of the twentieth century, especially in the West, the distress they suffered was intensified by psychiatrists and psychoanalysts who blamed the mother for the child's strange behaviour. The fact that there are no visible signs of the brain disorder underlying autism reinforced the tendency to look for a non-physical reason for the socially unacceptable behaviour. In some other societies, the shame of having a child with a developmental disorder adds immeasurably to the family's problems. One of the messages of this book is that personal experience of any illness or disability is deeply affected by the culture in which one lives.

Isabel's story is interwoven with the other themes. Her parents' acceptance and love for her is evident in everything Grinker writes. She has typically autistic visuospatial and memory skills, and a high IQ if the measurement is based on these abilities. But this is combined with the autistic lack of inborn social skills, so she has trouble adapting her abilities to social demands. It is strange to think how recently psychologists became aware that social skills are an important dimension of overall ability.

As a mother of a daughter with autism, I know that parents will empathize with the author and be fascinated by his view of the subject and the stories he tells. As a professional, I recommend this book to those working in the field for its thought-provoking examination of important themes. ∎

Lorna Wing is honorary consultant psychiatrist at the Centre for Social and Communication Disorders, 113 Masons Hill, Bromley, Kent BR2 9HT, UK.

Control without hierarchy

Understanding how particular natural systems operate without central control will reveal whether such systems share general properties.

Deborah M. Gordon

Because most of the dynamic systems that we design, from machines to governments, are based on hierarchical control, it is difficult to imagine a system in which the parts use only local information and the whole thing directs itself. To explain how biological systems operate without central control — embryos, brains and social-insect colonies are familiar examples — we often fall back on metaphors from our own products, such as blueprints and programmes. But these metaphors don't correspond to the way a living system works, with parts linked in regulatory networks that respond to environment and context.

Recently, ideas about complexity, self-organization, and emergence — when the whole is greater than the sum of its parts — have come into fashion as alternatives for metaphors of control. But such explanations offer only smoke and mirrors, functioning merely to provide names for what we can't explain; they elicit for me the same dissatisfaction I feel when a physicist says that a particle's behaviour is caused by the equivalence of two terms in an equation. Perhaps there can be a general theory of complex systems, but it is clear we don't have one yet.

A better route to understanding the dynamics of apparently self-organizing systems is to focus on the details of specific systems. This will reveal whether there are general laws. I study seed-eating ant colonies in the southwestern United States. In each ant colony, the queen is merely an egg-layer, not an authority figure, and no ant directs the behaviour of others. Thus the coordinated behaviour of colonies arises from the ways that workers use local information.

If you were the chief executive of an ant colony, you would never let it forage in the way that harvester ant colonies do. Put down a pile of delicious mixed bird-seed, right next to a foraging trail, and the ants will walk right over it on their way to search for buried shreds of seeds 10 metres further on. This behaviour makes sense only as the outcome of the network of interactions that regulates foraging behaviour.

Foraging begins early in the morning when a small group of patrollers leave the nest mound, meander around the foraging area and eventually return to the nest. A high rate of interactions with returning patrollers is what gets the foragers going, and through chemical signals the patrollers determine the foragers' direction of travel. Foragers tend to leave in the direction that the patrollers return from. If a patroller can leave and return safely, without getting blown away by heavy wind or eaten by a horned lizard, then so can a forager.

Once foraging begins, the number of

ants that are out foraging at any time is regulated by how quickly foragers come back with seeds. Each forager travels away from the nest with a stream of other foragers, then leaves the trail to search for food. When it finds a seed, it brings it directly back to the nest. The duration of a foraging trip depends largely on how long the forager has to search before it finds food. So the rate at which foragers bring food back to the nest is related to the availability of food that day. Foragers returning from successful trips stimulate others to leave the nest in search of food.

But why do foragers walk right past seed baits? We learned recently that during a day, each forager keeps returning to the same patch to search for seeds. Once a forager's destination for the day is set, apparently by the first find of the day, even a small mountain of seeds is not enough to change it. In this system, the success of a forager in one place, returning quickly to the nest with a seed, stimulates another forager to travel to a different place. A good day for

foraging in one place usually means a good day everywhere; for example, the morning after a heavy rain, seeds buried in the soil are exposed and can be found quickly.

The regulation of foraging in harvester ants does not use recruitment, in which some individuals lead others to a place with abundant food. Instead, without requiring any ant to assess anything or direct others, a decentralized system of interactions rapidly tunes the numbers foraging to current food availability.

It is difficult to resist the idea that general principles underlie non-hierarchical systems, such as ant colonies and brains. And because organizations without hierarchy are unfamiliar, broad analogies between systems are reassuring. But the hope that general principles will explain the regulation of all the diverse complex dynamical systems that we find in nature, can lead to ignoring anything that doesn't fit a pre-existing model.

When we learn more about the specifics of such systems, we will see where analogies between them are useful and where they break down. An ant colony can be compared to a neural network, but how do colonies and brains, both using interactions among parts that respond only to local stimuli, each solve their own distinct set of problems?

Life in all its forms is messy, surprising and complicated. Rather than look for perfect efficiency, or for another example of the same process observed elsewhere, we should ask how each system manages to work well enough, most of the time, that embryos become recognizable organisms, brains learn and remember, and ants cover the planet.

Deborah M. Gordon is in the Department of Biological Science, Stanford University, Stanford, California 94305-5020, USA.
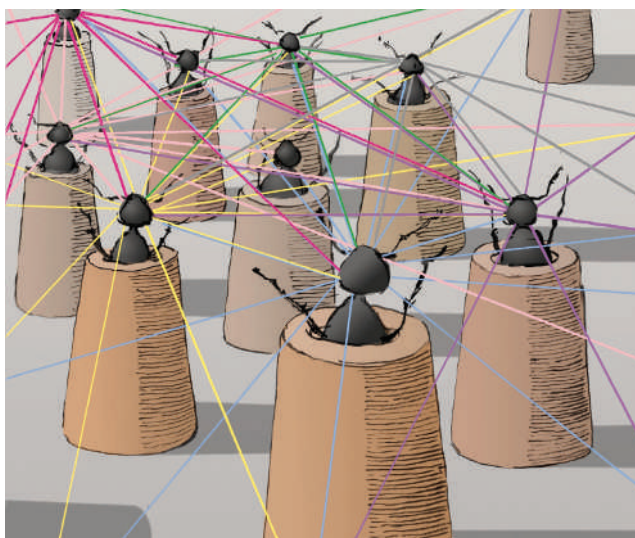
FURTHER READING
Gordon, D. M. *Ants at Work* (W. W. Norton and Co., New York, 2000).
Haraway, D. J. *Crystals, Fabrics, and Fields: Metaphors of Organicism in Twentieth-Century Developmental Biology* (Yale Univ. Press, New Haven, 1976).
Lewontin, R. C. *The Triple Helix: Gene, Organism and Environment* (Harvard Univ. Press, Cambridge, 2000).

For other essays in this series, see http:// nature.com/nature/focus/arts/connections/ index.html

J.KAPUSTA/IMAGES.COM

CONNECTIONS

NEWS & VIEWS

CANCER

Drivers and passengers

Daniel A. Haber and Jeff Settleman

Studies that have provided the first unbiased, large-scale analyses of DNA mutations across an array of cancers also have lessons for the proposal to annotate the entire cancer genome.

Cancer results from an accumulation of mutations and other heritable changes in susceptible cells. So far, abnormalities in about 350 genes have been implicated in human cancers, but the true number of 'cancer genes' is unknown. On page 153 of this issue, Greenman and colleagues[1] build on their previous analyses of breast, lung and brain tumours[2–4] by identifying mutations in the genes encoding all known protein kinases — enzymes that regulate other proteins through the addition of a phosphate residue — across various types of cancer. Together with a whole-genome resequencing analysis of a smaller number of breast and colorectal cancers published by Sjöblom *et al.*[5] in October 2006, this study presents a largely unbiased overview of the spectrum of mutations in human cancers.

Greenman and co-workers[1] undertook a comprehensive sequencing of 518 protein-kinase-encoding genes in 210 cancers. Kinases have been implicated in many aspects of tumorigenesis and several have now been validated as targets for drug therapy. The spectacular success of the drug imatinib (Gleevec) in treating chronic myeloid leukaemia stems from its suppression of a kinase known as BCR–ABL, which is the product of a gene located within a cancer-specific, translocated chromosome[6]. In their analysis of the collection of cellular kinases, the 'kinome', Greenman *et al.* identified 1,000 mutations. Mutations were relatively common in cancers of the lung, stomach, ovary, colon and kidney, and rare in cancers of the testis and breast, and in carcinoid tumours, which are usually found in the gastrointestinal tract. Tumours with defects in DNA-mismatch repair harboured large numbers of mutations, whereas other types
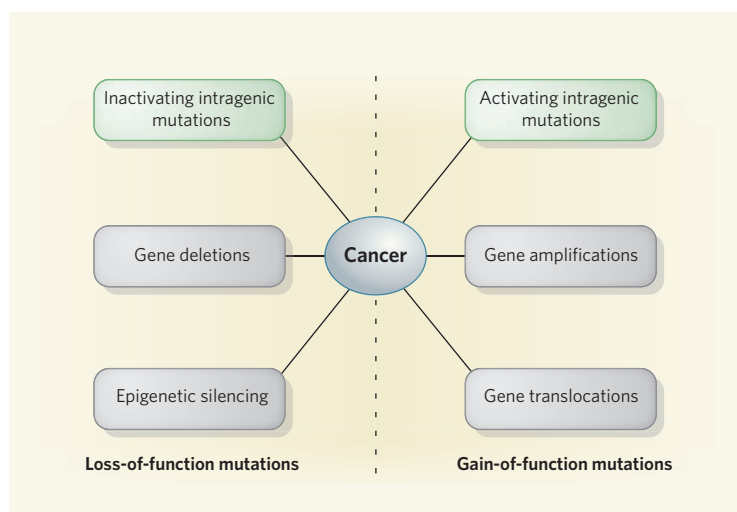

Figure 1 | The variety of defects underlying human cancer. DNA-sequencing strategies, such as those described by Greenman *et al.*[1] and Sjöblom *et al.*[5], are aimed at detecting small nucleotide changes within genes (intragenic mutations), and identify these genes as the target of a mutational event (highlighted in green). But various abnormalities may be transmitted from a cancer cell to its progeny, some of which activate specific genes (gain-of-function mutations), whereas others inactivate them (loss-of-function mutations). Alternative strategies, such as comparative genomic hybridization[12], are required to scan for deletion or amplification of chromosome fragments, which often contain many genes, making it more difficult to identify the specific gene(s) targeted by these events. Epigenetic silencing involves heritable modifications of nucleotides and histones in regulatory regions of genes, leading to suppression of gene expression in the absence of DNA mutations. Translocations lead to the fusion of DNA fragments from different chromosomal regions, either creating an abnormal fusion protein or leading to aberrant expression of a normal gene.

of tumour revealed no detectable mutations. Specific patterns of nucleotide substitution differed among cancers from various tissue types, possibly reflecting the effects of external mutagens or defects in DNA repair.

By the time a cancer is diagnosed, it comprises billions of cells carrying the DNA abnormalities that initiated malignant proliferation and many additional genetic lesions acquired along the way. Some of these secondary mutations emerge owing to selective pressure during tumorigenesis (drivers); others may be incidental (passengers), resulting from mutational exposures, genome instability or simply the large number of cell doublings that leads from a single transformed cell to a clinically detectable cancer. To distinguish driver from

passenger mutations, Greenman *et al.* used a statistical model comparing the observed-to-expected ratio of synonymous (no amino-acid change) mutations with that of non-synonymous (altered amino acid) mutations. An increased proportion of non-synonymous mutations implies selection pressure during tumorigenesis. Overall, 158 predicted driver mutations were identified in 120 genes encoding kinases. In contrast to the recurrent mutations in the gene encoding a kinase known as BRAF and previously identified by this group in malignant melanomas[7], most kinase mutations identified across different tumour types were 'single hits'.

Sjöblom and co-workers[5] used a different strategy, but reached similar conclusions. By initially sequencing about 13,000 genes in 11 breast and 11 colorectal cancer cell lines, they identified 1,307 validated nucleotide changes in 1,149 genes, of which 189 met their criteria for significance. Few overlapping driver mutations were identified between the kinase genes analysed in these two studies, highlighting the requirement for a very large number of samples to capture the full repertoire of genetic heterogeneity in cancer.

Indeed, the US National Cancer Institute and the National Human Genome Research Institute have proposed the Human Cancer Genome Project[8], which aims to sequence 12,500 tumour samples (250 specimens from 50 different cancers), focusing on at least 2,000 genes implicated in tumorigenesis. The scale of this proposed analysis is such that it should reveal recurrent mutations in subsets of cancer, an important clue to identifying truly significant drivers. Nonetheless, small nucleotide changes within genes that are detectable

by sequencing constitute only a subset of abnormalities underlying human cancer — gene amplification or deletion, inactivation of genes through epigenetic silencing, and chromosomal translocations also contribute to alterations in cancer genes (Fig. 1), and these would not be detectable by sequence analysis of known genes.

The Human Cancer Genome Project is likely to be costly at this time of crisis in US funding for biomedical research, and its ultimate value will depend on identifying crucial genetic lesions that point to more effective therapies from among the many driver mutations identified in such a screen. The exceptional dependence of chronic myeloid leukaemia on the translocated BCR-ABL kinase[6], and the dependence of some 10% of non-small-cell lung cancers on a mutated epidermal growth factor receptor kinase[9,10], are both correlated with dramatic responses to small-molecule inhibitors. These and, it is to be hoped, other genetically defined, highly drug-responsive subsets of cancer that are yet to be identified, exemplify the phenomenon of 'oncogene addiction', whereby only one of the many genetic lesions in a tumour proves to be its Achilles' heel[11]. Mutant kinases and other critically altered proteins in cancer cells may thus prove to be good drug targets; however, many other driver mutations that occur with tumour progression may not be essential for tumour maintenance. Therefore, complementary functional screens will be needed to validate the potential of suspected cancer genes as therapeutic targets.

Together, these initial sequencing studies[1,5] provide insight into mutational profiles in various cancers that might help to identify the molecular mechanisms responsible for tumour initiation and progression. But, in considering the application of sequencing strategies to a vast number of specimens from cancer patients, these studies also give a cautionary traffic signal — each cancer genome carries many unique abnormalities, and not all mutations identified contribute equally to the manifestation of the associated cancers. A combination of genetic and functional approaches will be essential to correctly identify true drivers from the many passengers on the road to tumorigenesis. ■

Daniel A. Haber and Jeff Settleman are at the Massachusetts General Hospital Cancer Center and Harvard Medical School, Building 149–7, Charlestown, Massachusetts 02129, USA.
e-mail: haber@helix.mgh.harvard.edu

1. Greenman, C. *et al. Nature* **446,** 153–158 (2007).
2. Stephens, P. *et al. Nature Genet.* **37,** 590–592 (2005).
3. Stephens, P. *et al. Nature* **431,** 525–526 (2004).
4. Hunter, C. *et al. Cancer Res.* **66,** 3987–3991 (2006).
5. Sjöblom, T. *et al. Science* **314,** 268–274 (2006).
6. Kantarjian, H. *et al. N. Engl. J. Med.* **346,** 645–652 (2002).
7. Davies, H. *et al. Nature* **417,** 949–954 (2002).
8. von Eschenbach, A. & Collins, F. http://cancergenome.nih.gov/about/TCGA_executive_summary.pdf (2005).
9. Lynch, T. J. *et al. N. Engl. J. Med.* **350,** 2129–2139 (2004).
10. Paez, J. G. *et al. Science* **304,** 1497–1500 (2004).
11. Weinstein, I. B. *Science* **297,** 63–64 (2002).
12. Pinkel, D. & Albertson, D. G. *Annu. Rev. Genomics Hum. Genet.* **6,** 331–354 (2005).

Silicon life forms

David J. Norris

A simple chemical reduction process has been used to replicate intricate natural networks of silica at a relatively low temperature. The equally elaborate product is made of silicon — electronics' golden boy.

Living organisms can produce amazingly complex materials — from bones to butterfly wings — that are not only functional but often beautiful[1,2]. A dream of materials scientists is to understand and mimic nature, and thereby to obtain better man-made materials[3,4]. This remains a thorny problem, and few technologically relevant materials have so far been obtained solely by mimicking biological processes; conventional fabrication steps are typically needed too.

On page 172 of this issue, Bao *et al.*[5] provide a powerful new tool for modifying biologically derived or inspired materials. They show how intricate glass skeletons, obtained from common algae, can be converted into silicon while maintaining their complex structure. Silicon is arguably the 'gold standard' among electronic materials, and this approach is akin to the magic touch of a modern Midas. It should allow a variety of intricate glass structures, both natural and artificial, to be transformed into silicon, and could have applications in sensing, electronics and optoelectronics.

Biological structures such as bones, shells and spines contain both inorganic and organic components. Nature uses these composites to achieve desirable mechanical properties in a material that can be formed under ambient conditions. The inorganic components tend to be limited to a few minerals such as phosphates, carbonates and silicates. The presence of silicates is perhaps not surprising: these materials combine oxygen and silicon, the two most abundant elements in Earth's crust. A common silicate is silica, SiO_2, which is found both in biological settings and in man-made materials such as cement and everyday glass. Silica has a rich chemistry that can be adapted to obtain a variety of structures[6].

With the notable exception of the optical fibres that are ubiquitous in telecommunications, silica oddly does not have a starring role in modern technology. In electronic circuits, for instance, the semiconductor silicon takes the lead, and silica is relegated to a supporting role of protecting, insulating and coating. But silica's abundance as a natural material is crucial in this context for another, often forgotten reason: the central player, silicon, is extracted from it. Heating silica at about 2,000 °C in the presence of carbon produces silicon. This can be processed further to obtain silicon wafers, which are among the purest available materials, and are the foundation for most of electronics.

Can silica be converted to silicon under milder conditions? Moreover, can this be achieved while preserving a complex shape? Bao *et al.*[5] provide a simple answer to these questions. They react silica with magnesium gas at 650 °C to obtain a solid that contains both silicon and magnesia (magnesium oxide) phases. Silicon is not very volatile at these tem-
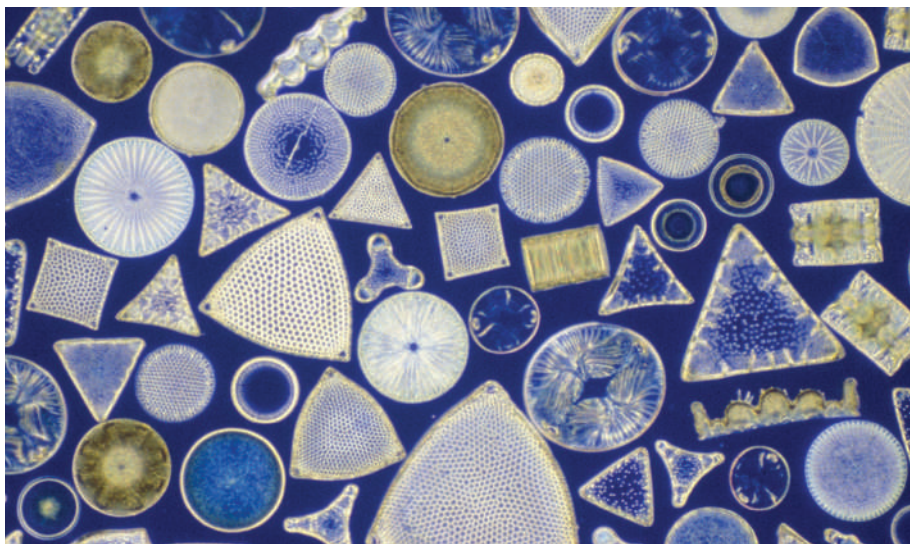

<div style="writing-mode: vertical">M. I. WALKER/SPL</div>

Figure 1 | Glass skeletons. Diatom species come in many different intricate forms. Bao and colleagues[5] show how the silica exoskeletons of diatoms can be converted into technologically relevant silicon while preserving their shape. The approach could also be used to convert the growing list of man-made nanostructured silica materials into silicon.

peratures and is partially trapped by the magnesia, with which it is intertwined, and so it does not move significantly during this reaction. Consequently, the original shape of the silica is largely preserved. The magnesia can be selectively removed by bathing the solid in hydrochloric acid, leaving behind a silicon replica of the original silica structure.

Bao and colleagues demonstrate their approach with several types of diatom. Because of their abundance, these unicellular, photosynthesizing algae are important in soil, freshwater and seawater ecosystems. Diatoms fortify their cell walls with silica, and so have also been used to study biomineralization[7]. Their exoskeletons exhibit a variety of beautifully intricate, species-specific shapes and patterns (Fig. 1).

The authors use their technique to convert these silica skeletons into silicon. Their results show that the overall shapes, with their intricate pores and channels, are conserved. Also, as oxygen is removed from the structure, fewer atoms remain. This introduces new, nanometre-scale pores into the replica. For some applications — those that require solid silicon — this would be a disadvantage. For others it can be desirable for two reasons. First, the presence of the pores significantly increases the surface area of the final material, which can be important for applications such as sensing. Second, the final structure is composed of nanoscale crystals of silicon, which can have useful optical properties. In particular, unlike bulk silicon, nanocrystalline silicon can fluoresce efficiently[8,9].

Bao et al.[5] test both of these effects. First, they attach wires to a single diatom replica and use it as a microsensor for nitric oxide. Their results suggest that the diatom-derived silicon structure can provide a much more efficient sensor element than other, more conventional approaches. Second, they find that their silicon replicas can fluoresce after being partially oxidized in water.

This work indicates that many different silicon materials can be derived from diatoms, as well as other silica structures harvested from nature. More generally, it complements other recently developed 'templating' techniques. Solids structured on nanometre or micrometre length scales can now be obtained through a variety of simple self-assembly routes. These templates are then filled with a different material of interest before the template is removed. Silica templates are common because they are so easy to prepare, and silicon has been used as an infill material[10] because of its technological importance and its high refractive index, which is useful in some applications in photonics.

A templating approach obviously produces the 'negative' of the original template. To obtain a duplicate of the template, the procedure must be repeated[11]: the negative structure is used as a mould, infiltrated a second time and removed. The mould material must be chosen carefully for all of these steps to be completed successfully. In contrast, Bao et al.[5] have provided a

method for copying a structure into silicon directly, albeit with the introduction of nanometre-scale pores. In combination with other self-assembly tricks, this opens the way to a variety of exciting new fabrication strategies. The implications might encompass not just biologically derived materials, but also many man-made creations. ∎

David J. Norris is in the Department of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Avenue SE, Minneapolis, Minnesota 55455, USA. He is currently on sabbatical at the Walter Schottky Institute, Garching, Germany.
e-mail: dnorris@umn.edu

1. Vukusic, P. & Sambles, J. R. Nature 424, 852–855 (2003).
2. Aizenberg, J. et al. Science 309, 275–278 (2005).
3. Heuer, A. H. et al. Science 255, 1098–1105 (1992).
4. Mann, S. & Ozin, G. A. Nature 382, 313–318 (1996).
5. Bao, Z. et al. Nature 446, 172–175 (2007).
6. Iler, R. K. The Chemistry of Silica (Wiley, New York, 1979).
7. Sumper, M. Science 295, 2430–2433 (2002).
8. Canham, L. T. Appl. Phys. Lett. 57, 1046–1048 (1990).
9. Littau, K. A., Szajowski, P. J., Muller, A. J., Kortan, A. R. & Brus, L. E. J. Phys. Chem. 97, 1224–1230 (1993).
10. Chomski, E. & Ozin, G. A. Adv. Mater. 12, 1071–1078 (2000).
11. Jiang, P., Bertone, J. F. & Colvin, V. L. Science 291, 453–457 (2001).

EVOLUTIONARY BIOLOGY

The Elvis paradox

Andrew Hendry

Evidence for a universal driver of evolution across all timescales could mean that the venerable paradox of stasis is dead. But even with such evidence, some biologists would be reluctant to accept its passing.

Disagreement has long swirled around the relative importance of various forces that might drive evolution on timescales ranging from dozens to millions of generations. Writing in *The American Naturalist*, Estes and Arnold[1] offer a provocative contribution to this debate: they propose that evolutionary changes on *all* timescales might be explained by a single, simple model of adaptation.

Much of the challenge can be distilled down to what has been called the 'paradox of stasis'[2]. For me, the most obvious manifestation of this paradox is that neo-darwinian theory, with its emphasis on the power of selection, predicts the potential for rapid adaptation, whereas most lineages of organisms instead show long-term stasis: that is, very little cumulative change over long periods of time[3,4]. Several hypotheses have been advanced in the hope of resolving this seeming discontinuity between short- and long-term evolution[2,4,5], but none has been convincing enough to resonate across the various camps.

Estes and Arnold[1] point out that the best way to discriminate between the hypotheses is to confront the predictions of alternative evolutionary models with the reality of data. This sort of comparison has recently been made possible by compilations of data on phenotypic changes (such as in mean body size) within animal lineages at a variety of different timescales[4,6,7]. The pattern emerging from these data is that phenotypic changes over dozens of generations can range from small to large, and that this range remains roughly the same even over millions of generations (Fig. 1). This pattern thus affirms the original paradox — that phenotypic change can be dramatic on short timescales, but rarely accumulates into substantial evolutionary trends.
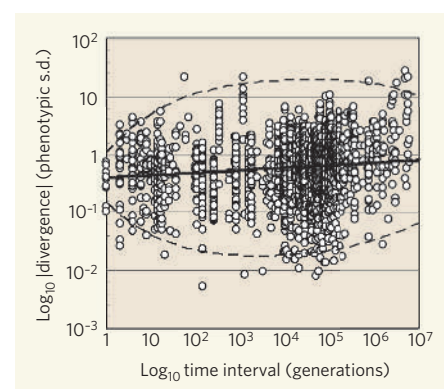

Figure 1 | The database of phenotypic changes analysed by Estes and Arnold[1]. The graphic summarizes phenotypic changes (usually body size or other aspects of morphology) through time within animal lineages, as well as phenotypic differences between lineages that had a common ancestor at a known time in the past. The lineages concerned include a wide variety of vertebrate and invertebrate taxa, ranging from snails to flies to salmon to mice to horses. Dotted lines are the 99% confidence ellipse for the data. The solid line is the regression line through the data. (Reproduced from ref. 1.)

Estes and Arnold[1] evaluate the degree to which six evolutionary models fit the observed data. All of these models are based on 'adaptive landscapes' — an analytical framework that relates mean phenotypes (mean body size, for instance) to the expected mean fitness (that is, number of offspring) of a population (Box 1, overleaf). Evolution on such landscapes tends towards 'hill climbing', where the mean phenotype of the population moves towards that which maximizes population fitness (a local fitness peak)[8,9].

Three of the models tested by Estes and

Box 1 | Evolution on six adaptive landscapes

Adaptive landscapes, shown here in blue, depict the mean fitness of individuals in a population (for example, average number of offspring) for a range of possible mean phenotypes (for example, average body size) in that population. A population with a particular phenotypic distribution (an example is shown in red) will tend to evolve up the steepest adjacent slope on the adaptive landscape until its mean approaches that which maximizes population fitness. Estes and Arnold[1] consider evolutionary changes that might result from several different types of adaptive landscape.

Under 'neutrality', the adaptive landscape is flat, without any fitness peaks, and so the mean phenotype will drift around at random (red arrows). Under

'brownian motion', the adaptive landscape has a single peak, the location of which drifts around at random (blue arrows). Under this model, the change in peak position is calculated from the position that the peak reached in the previous generation. Under 'white-noise motion', the peak again moves at random, but this time the change in each generation is calculated from an unchanging starting point (that is, the changes are not additive). For these models of random peak movement, mean phenotypes will follow the randomly moving peak.

Under 'moving optimum', the adaptive landscape has a single peak, and this peak is moving predictably in a particular direction every generation, with mean phenotypes following after. Under 'peak shift', the adaptive landscape has two fitness peaks separated
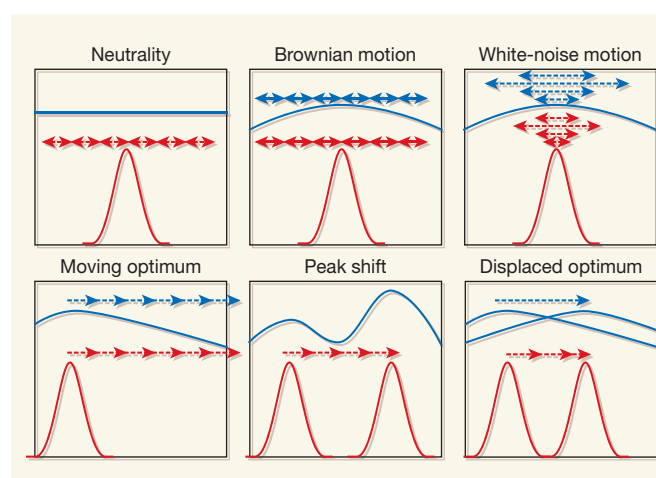
by a valley, and mean phenotypes may sometimes shift from the lower peak to the higher one. Under 'displaced optimum', the adaptive landscape has a single peak, and this peak jumps in a

single generation to a new location, and then stays there. The mean phenotype then evolves over to the new peak. It is this model that Estes and Arnold find best fits the data. **A.H.**

Arnold represent different flavours of randomness (Box 1). One model ('neutrality') evokes a flat adaptive landscape, on which the mean phenotype of a finite population will drift at random. The other two random models have a single fitness peak that moves randomly according to either 'brownian motion' or 'white-noise motion'. In these two models, mean phenotypes forever chase the randomly moving fitness peaks, like a new task for Sisyphus. Estes and Arnold argue that these three models fail to fit the data well, suggesting that randomness, at least in these forms, may not be a primary driver of phenotypic change. This conclusion will be reassuring, or perhaps just obvious, to the innumerable evolutionary biologists who believe that adaptation plays a central role in evolution.

The other three models involve a directional shift in the position of a fitness peak. In one ('moving optimum'), the adaptive landscape has a single peak that moves step-by-step in a particular direction, with the phenotypic mean of the population following along. This model predicts too little evolution on short timescales and too much on long timescales relative to the observed data. In another directional model ('peak shift'), the adaptive landscape has two peaks and, under some conditions, the phenotypic mean can shift rapidly from one peak to the other. This model predicts too little evolution on short timescales, and it only fits the data well on long timescales when populations are unrealistically small. In the final model ('displaced optimum'), the adaptive landscape has only one peak, and the position of this peak shifts abruptly, but just once. In the authors' estimation, this last model fits the data quite well.

The displaced-optimum model can be visualized by reference to a hypothetical population

that is well adapted to its local environment, in which case the mean phenotype of the population (let's say a body mass of 10 g) will be centred near a local fitness peak (also 10 g). Then imagine that the environment changes abruptly and displaces the optimum to a new location of 12 g, leaving the population mean behind at 10 g. Now the largest individuals in the population will be favoured by natural selection, and the mean phenotype will increase across generations until it is positioned at the new fitness peak (that is, 12 g), where it will then stay in the absence of further environmental change. This all makes sense, but the surprising part is that the displaced-optimum model assumes that this happens only once for a given lineage — regardless of timescale. The key general point, however, is that the peak can be displaced only a restricted amount (that is, within defined bounds), even if it takes several steps to get there.

Conveniently, Estes and Arnold[1] provide an Excel file in which every parameter in every model can be varied and the resulting outcomes compared with actual data. After playing with these models myself, I tend towards general agreement with the authors, adding the caveat that achieving the observed changes on short timescales requires a very large displacement of the optimum, coupled with a very sharp fitness peak. These properties mean that a population will have a substantial fitness decline immediately after the optimum moves — a possible recipe for extinction. Perhaps the largest changes on short timescales are the result of phenotypic plasticity (when a genotype expresses a different phenotype in a new environment), rather than of genetic change, the latter being the focus of the models. Other factors that may inflate short-term changes are sampling errors, conflation of

geographical variation with lineage evolution, and publication biases (perhaps published studies on short timescales tend towards those that find the largest changes).

Have Estes and Arnold[1] slain the paradox of stasis with a simple displaced-optimum model? In my opinion, the paradox might have been slain only in a broad sense, because the pattern in the data also might be replicated by other models that generate a range of short-term changes that do not accumulate into long-term trends. Such models might include various types of fluctuating selection, where fitness peaks on adaptive landscapes move back and forth owing to environmental variation[10]. Perhaps the paradox of stasis will have its final death at the point of a rapier, whereas Estes and Arnold have wielded a scimitar.

Whatever the model, it will have to generate rapid changes on short timescales, and yet still be constrained by boundaries on long timescales[4]. It is also possible that the paradox is a phantom, against which swords are of no use. Indeed, it may have been dead on arrival: way back in 1944, George Gaylord Simpson[8] suggested that evolutionary stasis might be explained by 'adaptive zones', where fitness peaks move back and forth within constrained bounds.

For some, any report of the death of this paradox will probably evoke the same reaction as the death of Elvis, with a large number of fans reluctant to accept its passing. But in the end, evolutionary biologists will probably converge on more pertinent questions, such as 'What generates and maintains adaptive zones in the first place?', and 'How do some lineages ultimately bridge the gap between different adaptive zones?'. This convergence would probably both please and frustrate Simpson, were he still alive, given that he posed much

50 YEARS AGO

Les Néanderthaliens. Par Étienne Patte — At the time of the original discovery in 1856, the Neanderthal skull aroused controversies which to-day seem to have been unnecessarily acrimonious, but then such controversies always do seem (and, because of the hidden emotions engendered, perhaps always will be) the fate of discoveries of early man or his progenitors... Perhaps more has been written and speculated about Neanderthal man than any other Palæolithic type, and there is no reason to suppose that the last word on his origin and fate has been pronounced... it is commonly accepted that he represents a separate species, *Homo neanderthalensis.* Whether this specific distinction is really warranted will probably only be determined with the accumulation of still more skeletal remains.   **W. E. Le Gros Clark**
From Nature 9 March 1957.

100 YEARS AGO

In a letter in NATURE (August 2, 1906) I gave an account of some experiments which I considered proved that the α particle as initially expelled is not charged... But it is clear that if, as Rutherford considers probable, the α particle carries a multiple charge, the results I published do not by themselves suffice... I had hoped long ere this to submit this point to an experimental test, which is simple enough to do by varying the strength of the field. But I very much regret I have no longer the essential facilities necessary to carry on the investigation, particularly the means of obtaining a steady supply of liquid-air, and there does not appear to be any immediate prospect of my being in a position to repeat the experiments. The question at issue is a somewhat fundamental one in the relations of electricity and matter... so nothing remains but to withdraw what I have already published.   **Frederick Soddy**
From Nature 7 March 1907.

the same questions more than 60 years ago. ■

Andrew Hendry is in the Redpath Museum and Department of Biology, McGill University, Montreal, Quebec H3A 2K6, Canada.
e-mail: andrew.hendry@mcgill.ca

1. Estes, S. & Arnold, S. J. *Am. Nat.* **169,** 227–244 (2007).
2. Hansen, T. F. & Houle, D. in *Phenotypic Integration* (eds Pigliucci, M. & Preston, K.) 130–150 (Oxford Univ. Press, 2004).
3. Gould, S. J. & Eldredge, N. *Paleobiology* **3,** 115–151 (1977).
4. Gingerich, P. D. *Genetica* **112–113,** 127–144 (2001).
5. Charlesworth, B. *et al. Evolution* **36,** 474–498 (1982).
6. Hendry, A. P. & Kinnison, M. T. *Evolution* **53,** 1637–1653 (1999).
7. Kinnison, M. T. & Hendry, A. P. *Genetica* **112–113,** 145–164 (2001).
8. Simpson, G. G. *The Tempo and Mode of Evolution* (Columbia Univ. Press, 1944).
9. Arnold, S. J. *et al. Genetica* **112–113,** 9–32 (2001).
10. Grant, P. R. & Grant, B. R. *Science* **296,** 707–711 (2002).

PLANETARY SCIENCE

Water cycling on Mars

Victor R. Baker

The Meridiani Planum region on Mars is rich in minerals derived from evaporation, but lacks a topography consistent with standing water. Do the deposits stem from upwelling groundwater early in the planet's history?

A succession of sophisticated spacecraft missions has led to spectacular advances in the understanding of Mars' global hydrology over the past few decades. One of many examples is the discovery of abundant hydrated sulphate salt minerals. These minerals are found at many locations on the planet — most notably at Meridiani Planum, the landing site of NASA's robotic Mars rover Opportunity — and prove that water must once have been abundant on the surface of Mars. On page 163 of this issue, Andrews-Hanna *et al.*[1] use a numerical model to simulate the evolving global flow of sub-surface groundwater early in Mars' geological history. They place their simulation in the context of the formation of the enormous volcanic uplift feature known as Tharsis.

One way of developing a model of martian hydrology comes from a comparison with what we know about Earth. Western science was painfully slow in achieving its understanding of Earth's hydrological cycle. Many, if not most, of Isaac Newton's scientific contemporaries held the view that Earth's rivers were ultimately fed from upland springs. The springs were presumed to discharge water from within the planet, in much the same way as blood flows from a cut in an artery of the human body. Water from the oceans was presumed to return to the land through subsurface veins.

By contrast, Eastern philosophical writings had long held that Earth's water flowed as

Figure 1 | A well run dry? The Burns formation of Meridiani Planum is rich in evaporites. It is one of the sites where groundwater came to the surface and evaporated early in Mars' history, according to a model of the planet's hydrography produced by Andrews-Hanna *et al.*[1]. Earlier evidence from the Mars rover Opportunity for past evaporation of liquid water at the site had been difficult to explain in the absence of a topographic basin where water could stand. In the right foreground lies the Wopmay rock, whose distinctive lumpy appearance could be due to cracking processes caused by exposure to water.

NASA/JPL/CORNELL

part of a great cycle involving the atmosphere. About 3,000 years ago, the Vedas, Hindu texts of ancient India, explained Earth's water movements in terms of cyclical processes of evaporation, condensation, cloud formation, rainfall, river flow and water storage[2]. This concept of a global water cycle entered Western thought only late in the seventeenth century, when Edmond Halley, among others, showed that the evaporation from Earth's oceans supplied the rain clouds that led to a balancing run-off of water from land to the seas. Specifically, Halley compared evaporation from the Mediterranean Sea with estimates of its river inflow, thereby providing a modern scientific expression of the hydrological cycle.

Earth is a relatively warm and wet planet with a substantial atmosphere. So how does its water cycle relate to that of Mars, with that planet's extremely cold and dry climate and tenuous atmosphere? Although the conditions on Mars' present-day surface do not favour the persistence of exposed surface water or ice, high-resolution imagery of the planet has revealed distinctive, very ancient landforms typical of erosion by water and ice occurring at the planet's surface[3]. The evidence from surface features, such as fluvial valley networks and giant flood channels, strongly suggests that, early in its geological history, Mars was a 'water planet' like Earth[4]. Nevertheless, this hypothesis was not widely accepted until 2002, when NASA's Mars Odyssey mission revealed abundant subsurface ice on the planet[5].

Since then, products of the alteration of clay minerals by water[6] and evaporite salts[7] — minerals left behind when water evaporates — have been identified geochemically, further confirming that the early history of Mars involved liquid water. Most spectacularly, the Burns formation, a 7-metre-thick exposure analysed by the rover Opportunity, provides a clear geological record of evaporation in an arid surface environment[8] (Fig. 1). The chemistry and mineralogy of the deposit imply a fluctuating water table with acidic groundwater[9]. Thus, shallow, temporary, evaporating pools would have formed on the planet's surface, in much the same manner as they form under similar conditions on Earth at White Sands in the New Mexico desert[8]. But the Meridiani Planum region, where Opportunity detected mineral deposits from evaporation, is a tilted plain lacking topographic basins in which large bodies of water could stand and evaporate. How, then, did water get to this region of Mars?

Andrews-Hanna et al.[1] find that the Burns formation of Meridiani Planum is in fact exactly where a global groundwater simulation model for early Mars predicts upwelling of water on the ancient planet's surface. The picture is not dissimilar to the early ideas about Earth's subsurface veins of water that preceded Edmond Halley's investigations. Indeed, pictures from the new High Resolution Imaging Experiment on NASA's Mars Reconnaissance Orbiter show rocks that were once below the surface and have

been altered through the influence of flowing water[10]. This evidence of subsurface water flow was found along fractures in Mars' low-lying sedimentary deposits, thereby confirming an ancient groundwater circulation.

Andrews-Hanna and colleagues suggest that, in its early history, Mars had a globally connected groundwater system. The water would have circulated for an extended period of time, leaching solutes from the subsurface rocks. In this model, the water table reaches the surface at locations where the surface slopes down towards the northern lowlands of Mars. This occurs, for example, in the Meridiani Planum region, which lies just south of Mars' equator. In this way, groundwater would have been supplied continuously from the subsurface, its chemical evolution reflected in the dissolved salts. When the water evaporated, the salt would have been deposited at the planet's surface. The process occurred on an immense scale, deriving subsurface flow from large areas of Mars' highlands.

Critical to the long-term evolution of subsurface water flow on Mars was the development of the Tharsis rise. Early in martian geological history, this huge volcanic bulge developed on the planet's surface through immense outpourings of basaltic lava that also injected prodigious amounts of water and carbon dioxide into the young planet's atmosphere[11]. The huge pile of volcanic rocks at Tharsis progressively warped the surface of the planet and, as the authors show[1], thereby altered the groundwater flow. Precipitation from the very early Earth-like atmosphere recharged groundwater in the higher regions. The water then moved through

underground aquifers to discharge points such as Meridiani Planum.

The simulations by Andrews-Hanna et al.[1] solve the mystery of evaporite minerals in the absence of a topographic basin that would facilitate evaporation, and they help us to understand the role of topographic changes in Mars' ancient water flow. A final note of interest is that emerging subsurface acidic waters on Earth, such as those at Rio Tinto in southwestern Spain, are host to diverse living microorganisms[12]. Could Mars similarly have had a subsurface biosphere of microorganisms, which might have retreated to their groundwater refuge when water failed to emerge on the planet's surface during its later geological history? ■

Victor R. Baker is in the Department of Planetary Sciences and the Department of Hydrology and Water Resources, University of Arizona, Tucson, Arizona 85721, USA.
e-mail: baker@hwr.arizona.edu

1. Andrews-Hanna, J. C., Phillips, R. J. & Zuber, M. T. Nature 446, 163–166 (2007).
2. Chandra, S. Hydrology in Ancient India (Natl Inst. Hydrol., Roorkee, India, 1990).
3. Baker, V. R. Nature 412, 228–236 (2001).
4. Baker, V. R. The Channels of Mars (Univ. Texas Press, Austin, 1982).
5. Boynton, W. V. et al. Science 296, 81–85 (2002).
6. Bibring, J.-P. et al. Science 312, 400–404 (2006).
7. Squyres, S. W. et al. Science 313, 1403–1407 (2006).
8. Grotzinger, J. P. et al. Earth Planet. Sci. Lett. 240, 11–72 (2005).
9. McLennan, S. M. et al. Earth Planet. Sci. Lett. 240, 95–121 (2005).
10. Okubo, C. H. & McEwen, A. S. Science 315, 983–985 (2007).
11. Jakosky, B. M. & Phillips, R. J. Nature 412, 237–244 (2001).
12. Fernández-Remolar, D. C. et al. Earth Planet. Sci. Lett. 240, 149–167 (2005).

PLANT BIOLOGY

The force from without

Ben Scheres

It seems that the epidermis is the cell layer through which growth-promoting plant hormones called brassinosteroids exert their effect on cell expansion — a finding that puts a new perspective on classical views of plant growth.

Brassinosteroids are growth-promoting plant hormones, and their absence, or the absence of their main transmembrane receptor protein, creates dwarfed plants. In the past few years, both the brassinosteroid biosynthetic pathway and the molecular signalling route from the brassinosteroid receptor to intracellular gene-transcription factors have been elucidated[1,2]. That information, and the ability to restore normal growth selectively in specific plant cell layers, has now allowed Savaldi-Goldstein and co-workers (page 199 of this issue)[3] to clarify the pivotal role of a single cell layer — the outermost layer, known as the epidermis — in plant growth.

One of the main differences between

animals and plants is that animal cells can move relative to each other, whereas plant cells are practically glued together by cell walls. Thus, whereas animals can build beautiful structures through the slipping and sliding of cell sheets across each other, plants must shape themselves solely by the coordinated division and expansion of individual cells. The intimate connectedness of plant cells through rigid walls creates a mechanical network that can be driven or constrained by individual cell layers — just as you can stretch bubble gum by blowing it up or by tearing it apart with your fingers. This notion that certain tissue layers can act as masters of plant growth, whereas others follow like slaves, has been around in plant biology for a long
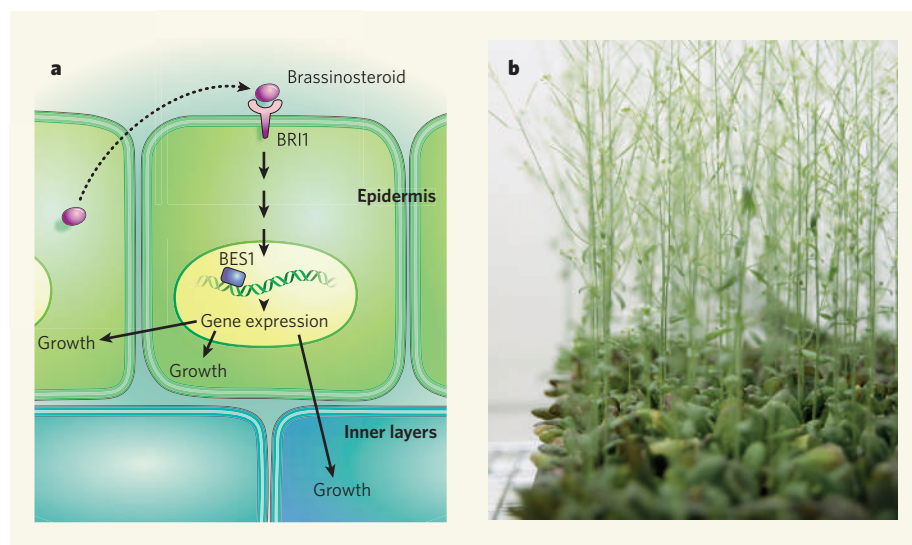
Figure 1 | Control of plant organ growth by signalling in the epidermis. a, The plant hormone brassinosteroid binds to its plasma membrane receptor BRI1, resulting in the activation of gene-transcription factors such as BES1, and thus modified gene expression[2]. Savaldi-Goldstein et al.[3] now show that these signalling events promote growth not only in the epidermal layer, but also in the inner cell layers. Consequently, they suggest that the epidermal layer controls growth throughout the leaves and other shoot-derived organs of flowering plants such as *Arabidopsis thaliana* (**b**).

time, and it has been proposed that the epidermis drives growth 'from without'[4,5] (Fig. 1a).

Savaldi-Goldstein and colleagues[3] set out to study the contribution of the epidermis to shoot growth. They used a mutant of the flowering plant *Arabidopsis thaliana* (Fig. 1b) that lacks both copies of the *CPD* gene. This gene is involved in brassinosteroid biosynthesis and the plant is therefore dwarfed. The authors found that by supplying the epidermal layer of this mutant with an intact version of the *CPD* gene, tagged with a fluorescent 'reporter' molecule, dwarfism could be reversed. When this gene was expressed in the inner tissue layers, it was far less effective in overcoming the growth defect.

Furthermore, dwarfism in plants that had mutations in the *bri1* gene, which encodes the brassinosteroid receptor BRI1, could be reversed by expressing this gene in the epidermal layer only. This confirmed that the receptor-binding ability of brassinosteroids in the epidermis is important for growth. Nonetheless, although growth recovered in these plants, certain aspects of inner-tissue differentiation that also require brassinosteroid signalling — more specifically, the organization of vascular tissues — did not. Moreover, expression of the *bri1* gene only in the vascular layer could not overcome the growth defect.

In an experiment that combined reactivation both of brassinosteroid biosynthesis in the inner tissues and of its receptor-binding ability in the epidermis, the authors confirmed that the inner tissues made only a small contribution to growth. These data convinced Savaldi-Goldstein et al. that the brassinosteroid-mediated pathway in the epidermal tissue can drive growth. But they wondered whether the epidermis was also the tissue responsible for limiting general organ growth under normal conditions. They thus expressed in the various tissue layers a fluorescently tagged enzyme that breaks down brassinosteroids. Consistent with the idea that the epidermis is of major importance for growth, reduced plant stature was particularly obvious when the enzyme was restricted to the epidermis.

These experiments indicate that growth of the epidermal layer in a shoot is crucial for controlling total plant growth. This means that the epidermis must signal to the underlying tissues to keep their growth in step and prevent the plant from being torn apart. The authors therefore investigated how the epidermis might promote the growth of underlying tissues. Plant gene-transcription factors can move through intracellular channels called plasmodesmata. Growth-promoting gene-transcription factors downstream of brassinosteroid signalling have been identified[6], and Savaldi-Golstein et al. showed that tagged versions of these factors could overcome the dwarfism associated with mutations in the *bri1* gene when expressed in the epidermis. One of these transcription factors (BES1) was able to drive even greater growth than was seen in normal plants.

The authors' experiments using fluorescent tags revealed that these transcription factors relay their growth-promoting effect to the inner tissues without moving from the epidermis; thus, the nature of the growth-promoting signal between the epidermis and the inner layers remains unknown. Is the mobile signal chemical in nature? Or is it mechanical, with the epidermis pulling the inner tissues along? One way of addressing this issue is to elucidate the 'growth machinery' regulated by these hormones. Target-gene analysis of transcription factors such as BES1 should identify which candidates regulate cell expansion in the epidermis and whether these epidermal growth signals influence inner tissues.

Many other questions remain. Do the inner tissues respond only by modulating cell expansion? How does the brassinosteroid growth-control pathway fit in with other mechanisms of growth regulation, for example the regulation of growth repressors such as members of the DELLA family, by various plant growth substances[7]? Is the brassinosteroid response one pathway among several that affect epidermal growth, as is perhaps suggested by the limited overlap in target genes for the different growth-control pathways[8,9]? Do the various pathways target different layers? And finally, how does an essential role of the epidermis in growth, as observed by Savaldi-Goldstein and colleagues[3], relate to genetic analyses in developing flowers, which suggest that the inner tissue layers dictate the identity and thus the development and growth of organs[10–12]? Perhaps the control of growth involves sequential communication pathways in which the dominance of a layer in one phase of development is followed by dominance of another layer at a later stage. The brassinosteroid pathway dissected by Savaldi-Goldstein et al., together with newly available molecular tools such as fluorescently tagged, functional signal-transduction components, are excellent means to address these questions, and to further our understanding of the integrated control of plant growth. ∎

Ben Scheres is in the Department of Biology, Science Faculty, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands.
e-mail: b.scheres@bio.uu.nl

1. Fujioka, S. & Yokota, T. *Annu. Rev. Plant Biol.* **54,** 137–164 (2003).
2. Li, J. *Curr. Opin. Plant Biol.* **8,** 526–531 (2005).
3. Savaldi-Goldstein, S., Peto, C. & Chory, J. *Nature* **446,** 199–202 (2007).
4. Sachs, J. *Handbuch der Experimentalphysiologie der Pflanzen* (Engelmann, Leipzig, 1865).
5. Van Overbeek, J. & Went, F. W. *Bot. Gaz.* **99,** 22–41 (1937).
6. Yin, Y. *et al. Cell* **120,** 249–259 (2005).
7. Achard, P. *et al. Science* **311,** 91–94 (2006).
8. Nemhauser, J. L. *et al. PLoS Biol.* **2,** e258 (2004).
9. Nemhauser, J. L. *et al. Cell* **126,** 467–475 (2006).
10. Szymkowiak, E. J. & Sussex, I. M. *Plant Cell* **4,** 1089–1100 (1992).
11. Vincent, C. A., Carpenter, R. & Coen, E. S. *Plant J.* **33,** 765–774 (2003).
12. Sieburth, L. E. *et al. Development* **125,** 4303–4312 (1998).

Correction

In the News & Views article "Organic chemistry: A tuxedo for iodine atoms" by Phil S. Baran and Thomas J. Maimone (*Nature* **445,** 826–827; 2006), a couple of errors crept into the accompanying figure. The bond to the top-right hydrogen atom in hopene, which was indicated to be projecting below the plane of the paper, should be pointing upwards; in other words, the hashed bond to this atom should be replaced with a wedge-shaped bond. Similarly, the wedge-shaped bond at the bottom right of the promoter should be replaced with a hashed bond, to indicate that it is projecting below the page, and not above it.

SCANNING-TUNNELLING SPECTROSCOPY

Scanning-tunnelling spectra of cuprates

Arising from: J. Lee et al. Nature 442, 546–550 (2006)

The study of bosonic modes that couple to the charge carriers is a key element in understanding superconductivity. Using atomic-resolution scanning-tunnelling microscopy (STM) to extract the spectrum of these modes in the high-temperature superconductor $Bi_2Sr_2CaCu_2O_{8+\delta}$, Lee et al.[1] find a mode whose frequency does not depend on doping but that changes on isotopic substitution of ^{16}O with ^{18}O. From this, they infer a role for lattice modes (phonons). However, examination of their data reveals a weaker, but distinct, feature that has all the characteristics of the magnetic excitation identified as the bosonic mode in other competing experiments[2–4]. We therefore suggest that the lattice mode seen by Lee et al.[1] is not relevant to superconductivity and is due to inelastic tunnelling through the insulating oxide layer[5].

STM spectroscopy has been used in conventional low-temperature superconductors to extract a spectrum that agrees in detail with phonons found in the same materials by neutron spectroscopy[6]. It has also been applied to high-temperature superconductors by Zasadzinsky et al.[2,3] using superconductor–insulator–superconductor (SIS) conductance in break junctions. These authors found a sharp bosonic peak whose spectroscopic properties agreed with those of the well known neutron resonance: the frequency of the mode was proportional to the superconducting transition temperature, and the mode strength decreased markedly with doping. Similar doping dependence of the mode was seen by Hwang et al. using optical spectroscopy[4]. Scanning-tunnelling superconductor–insulator–normal-metal (SIN) spectroscopy gives a direct measure of the density of states of the superconducting state, whereas SIS break-junction tunnelling spectroscopy produces a convolution of the two superconducting densities of states that can be analysed to yield a bosonic spectrum.

A plausible explanation for the observations of Lee et al. is offered by Pilgram et al.[5], who suggest that the phonon found by Lee et al. was the result of the direct excitation of the apical oxygen vibrations, and not density-of-state effects associated with the superconducting CuO_2 layer. We propose a mechanism that reconciles the two sets of tunnelling experiments, as well as optics experiments, with the suggestion of Pilgram et al.[5]. The superconductivity gives a negative peak in the second derivative of the tunnelling current[6] and a positive one for the inelastic tunnelling mechanism[7]. Although Lee et al. observe a positive peak, this does not rule out a contribution from the superconducting plane because there is an additional positive structure in SIN tunnelling on the high-energy side of the main negative peak[6].

Figure 4b of Lee et al.[1] shows that the histograms of the maxima of d^2I/dV^2 are asymmetrical and consist of two components: a sharp, narrow, doping-independent peak and a second, broad component whose position and strength vary with doping. The procedure used by the authors to plot these curves would be expected to lose all spectroscopic information about the bosonic spectral function apart from the frequency of the maximum point. However, the doping-dependent asymmetry in their Fig. 4b shows that this is not the case.

There are several mechanisms that might reveal weaker spectral features in the histograms: for example, weaker features will be selected from time to time through dithering, in the presence of sufficient noise; or surface inhomogeneities might favour one process over the other in different parts of the sample. As shown in Fig. 1, the data of Lee et al.[1] are well fitted by a sum of two gaussian curves, one centred at 52 meV and the other with a variable-centre frequency. The amplitude of the broad peak decreases with

doping, and the centre frequency follows the superconductor's critical temperature (T_c), being lowest for the highly underdoped and highly overdoped samples; Fig. 1f shows the position of the broad component, normalized to the superconducting energy gap Δ, plotted against Δ, together with tunnelling data from Zasadzinski et al.[2], with which it agrees. These results are consistent with a magnetic-resonance model for the peak[2].

It is interesting that the broad component is very weak for the most overdoped sample. This is consistent with the proposal by Hwang et al., based on optics data[4], that the magnetic-resonance contribution to the self-energy weakens in the overdoped region for $T_c < 50$ K, in qualitative agreement with SIS tunnelling. Also, the reason that the spectra of Zasadzinski et al.[2,3] and Hwang et al.[4] do not show the phonon contribution is that they probe the properties of the conducting CuO_2 plane, whereas — in the light of the proposal by Pilgram et al.[5] — STM spectra contain a contribution from inelastic tun-
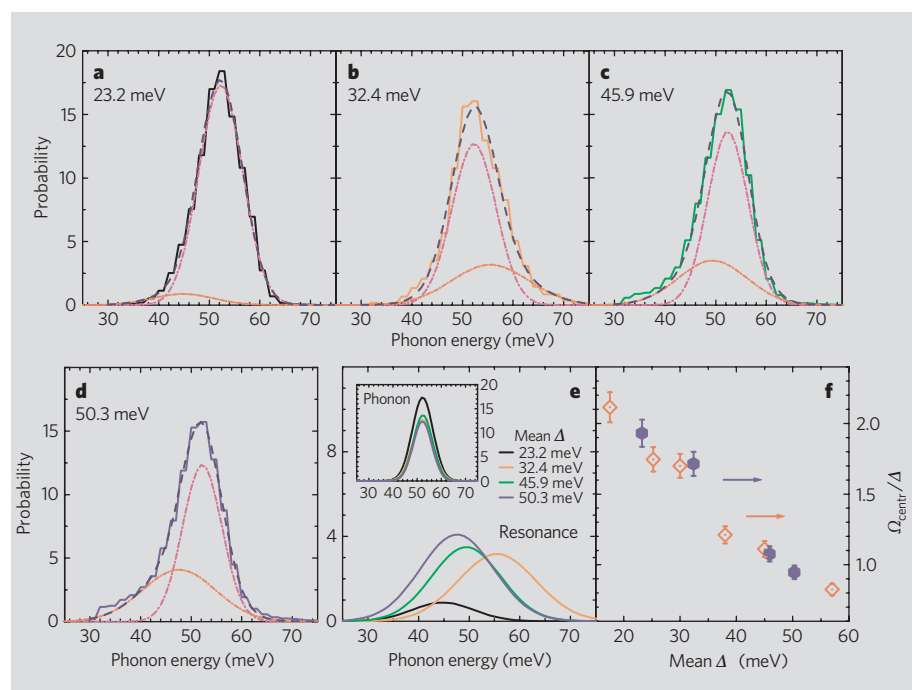

Figure 1 | Magnetic and lattice bosonic modes in STM spectra of cuprates. a–d, Bosonic peaks, reproduced from Lee et al.[1], for samples with different degrees of doping (full curves), from the most overdoped (**a**, black line) to the most underdoped (**d**, blue line), labelled with the corresponding value of the superconducting gap (in meV). Dashed curves, least-squares fit to a two-component model, a sharp peak fixed at a frequency of 52 meV (pink line), and a broader component that was allowed to vary in frequency and width (orange line). **e,** Plot showing the two components of the fit — the broad component (resonance) and the sharp mode (phonon) in the inset; **f,** plot comparing the centre frequency of the broad component (purple hexagons), normalized to the superconducting energy gap (Δ), with data from break-junction tunnelling from ref. 2 (orange diamonds). For a better comparison, we used the bias voltage that gives the maximum positive slope in the peak–dip–hump structure of the tunnelling conductance as Ω_{centr}, in which Ω is the phonon energy, instead of the bias voltage at the dip. J. Lee et al. generously shared some of their data with us.

nelling through the SrO layer. Our analysis shows that the STM technique can also see the bosonic spectrum that controls the self-energy of superconducting carriers and might therefore lead to the identification of magnetic fluctuations, such as the 'pairing glue' that is responsible for high-temperature superconductivity[8].

Jungseek Hwang*, Thomas Timusk*†, Jules P. Carbotte*†

*Department of Physics and Astronomy, McMaster University, Hamilton, Ontario L8S 4M1, Canada
e-mail: timusk@mcmaster.ca
†Canadian Institute of Advanced Research, Toronto, Ontario M5G 1Z8, Canada

1. Lee, J. *et al.* *Nature* **442,** 546–550 (2006).
2. Zasadzinski, J. F. *et al.* *Phys. Rev. Lett.* **87,** 067005 (2001).
3. Zasadzinski, J. F. *et al.* *Phys. Rev. Lett.* **96,** 017004 (2006).
4. Hwang, J., Timusk, T. & Gu, G. D. *Nature* **427,** 714–717 (2004).
5. Pilgram, S., Rice, T. M. & Sigrist, M. *Phys. Rev. Lett.* **97,** 117003 (2006).
6. McMillan, W. L. & Rowell, J. M. *Phys. Rev. Lett.* **14,** 108–112 (1965).
7. Jaklevic, R. C. & Lambe, J. *Phys. Rev. Lett.* **17,** 1139–1140 (1966).
8. Scalapino, D. J. *Nature Phys.* **2,** 593–594 (2006).

ARTICLES

Patterns of somatic mutation in human cancer genomes

Christopher Greenman[1], Philip Stephens[1], Raffaella Smith[1], Gillian L. Dalgliesh[1], Christopher Hunter[1], Graham Bignell[1], Helen Davies[1], Jon Teague[1], Adam Butler[1], Claire Stevens[1], Sarah Edkins[1], Sarah O'Meara[1], Imre Vastrik[2], Esther E. Schmidt[2], Tim Avis[1], Syd Barthorpe[1], Gurpreet Bhamra[1], Gemma Buck[1], Bhudipa Choudhury[1], Jody Clements[1], Jennifer Cole[1], Ed Dicks[1], Simon Forbes[1], Kris Gray[1], Kelly Halliday[1], Rachel Harrison[1], Katy Hills[1], Jon Hinton[1], Andy Jenkinson[1], David Jones[1], Andy Menzies[1], Tatiana Mironenko[1], Janet Perry[1], Keiran Raine[1], Dave Richardson[1], Rebecca Shepherd[1], Alexandra Small[1], Calli Tofts[1], Jennifer Varian[1], Tony Webb[1], Sofie West[1], Sara Widaa[1], Andy Yates[1], Daniel P. Cahill[3], David N. Louis[3], Peter Goldstraw[4], Andrew G. Nicholson[4], Francis Brasseur[5], Leendert Looijenga[6], Barbara L. Weber[7], Yoke-Eng Chiew[8], Anna deFazio[8], Mel F. Greaves[9], Anthony R. Green[10], Peter Campbell[1], Ewan Birney[2], Douglas F. Easton[11], Georgia Chenevix-Trench[12], Min-Han Tan[13], Sok Kean Khoo[13], Bin Tean Teh[13], Siu Tsan Yuen[14], Suet Yi Leung[14], Richard Wooster[1], P. Andrew Futreal[1] & Michael R. Stratton[1,9]

Cancers arise owing to mutations in a subset of genes that confer growth advantage. The availability of the human genome sequence led us to propose that systematic resequencing of cancer genomes for mutations would lead to the discovery of many additional cancer genes. Here we report more than 1,000 somatic mutations found in 274 megabases (Mb) of DNA corresponding to the coding exons of 518 protein kinase genes in 210 diverse human cancers. There was substantial variation in the number and pattern of mutations in individual cancers reflecting different exposures, DNA repair defects and cellular origins. Most somatic mutations are likely to be 'passengers' that do not contribute to oncogenesis. However, there was evidence for 'driver' mutations contributing to the development of the cancers studied in approximately 120 genes. Systematic sequencing of cancer genomes therefore reveals the evolutionary diversity of cancers and implicates a larger repertoire of cancer genes than previously anticipated.

Cancers are clonal proliferations that arise owing to mutations that confer selective growth advantage on cells. The mutated genes that are causally implicated in cancer development are known as 'cancer genes' and more than 350 have thus far been identified (ref. 1 and http://www.sanger.ac.uk/genetics/CGP/Census/). Cancer genes have been identified by several different physical and genetic mapping strategies, by biological assays and as plausible biological candidates. Each of these approaches has identified a subset of cancer genes, leaving the possibility that others have been overlooked. The provision of the human genome sequence, therefore, led to the proposal that systematic resequencing of cancer genomes could reveal the full compendium of mutations in individual cancers and hence identify many of the remaining cancer genes[2].

Somatic mutations occur in the genomes of all dividing cells, both normal and neoplastic. They may occur as a result of misincorporation during DNA replication or through exposure to exogenous or endogenous mutagens. Cancer genomes carry two biological classes of somatic mutation arising from these various processes. 'Driver' mutations confer growth advantage on the cell in which they occur,

are causally implicated in cancer development and have therefore been positively selected. By definition, these mutations are in 'cancer genes'. Conversely, 'passenger' mutations have not been subject to selection. They were present in the cell that was the progenitor of the final clonal expansion of the cancer, are biologically neutral and do not confer growth advantage. A challenge to all systematic mutation screens will, therefore, be to distinguish driver from passenger mutations. However, the prevalence and characteristics of driver and passenger mutations in cancer genomes are not currently well defined. The aim of these studies was to survey the numbers and patterns of somatic point mutations in a diverse set of human cancer genomes and hence to obtain insights into the relative contributions of driver and passenger mutations.

Somatic protein kinase mutations

The protein kinase gene family was selected for these studies because the protein kinase is the domain most commonly found among known cancer genes[1] and because inhibitors of mutated protein kinases have recently shown remarkable efficacy in cancer treatment[3].

[1]Cancer Genome Project, Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. [2]EMBL-European Bioinformatics Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK. [3]Molecular Pathology Unit, Neurosurgical Service and Center for Cancer Research, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114, USA. [4]Royal Brompton Hospital, London SW3 6NP, UK. [5]Ludwig Institute for Cancer Research, 1200 Brussels, Belgium. [6]Laboratory of Pathology/Experimental Patho-Oncology, Erasmus MC University Medical Center Rotterdam, Daniel den Hoed Cancer Center, Josephine Nefkens Institute, 3000 DR Rotterdam, UCL 745, B-1200, The Netherlands. [7]University of Pennsylvania Cancer Centre, Philadelphia, Pennsylvania 19104-6160, USA. [8]Department of Gynaecological Oncology, Westmead Hospital and Westmead Institute for Cancer Research, University of Sydney at the Westmead Millennium Institute, Westmead NSW 2145, Australia. [9]Institute of Cancer Research, Sutton, Surrey SM2 5NG, UK. [10]Department of Haematology, Addenbrooke's NHS Trust and University of Cambridge, Cambridge CB2 0QQ, UK. [11]Cancer Research UK Genetic Epidemiology Unit, University of Cambridge, Cambridge CB1 8RN, UK. [12]Queensland Institute of Medical Research, Royal Brisbane Hospital, Herston, Queensland 4029, Australia. [13]Van Andel Research Institute, Grand Rapids, Michigan 49503, USA. [14]Department of Pathology, The University of Hong Kong, Queen Mary Hospital, Pokfulam Road, Hong Kong.

Furthermore, the coding sequences of the protein kinases (Supplementary Table 3) constitute a much larger sample of cancer genome, approximately 1.3 Mb of DNA per case, than has previously been analysed across many cancer types, thus permitting insights into the general patterns of somatic mutation in human cancers.

Human cancers ($n = 210$) including breast, lung, colorectal, gastric, testis, ovarian, renal, melanoma, glioma and acute lymphoblastic leukaemia (Supplementary Table 3) were screened for somatic mutations in the coding exons and splice junctions of the 518 protein kinase genes[4]; a total of 274 Mb of cancer genome. Of the 210 cancers analysed 169 were primary tumours, 2 were early cultures and 39 were immortal cancer cell lines.

One-thousand-and-seven somatic mutations were detected (Supplementary Table 2 and http://www.sanger.ac.uk/genetics/CGP/Studies/). Of these, 921 were single base substitutions, 78 were small insertions or deletions and 8 were complex changes, usually double nucleotide substitutions. Of the single base substitutions, 620 encoded mis-sense changes, 54 caused nonsense changes, 28 were at highly conserved positions of splice junctions and 219 were synonymous (silent) mutations. Approximately one-third of these mutations have previously been reported[5–8].

Prevalence of somatic mutations

Although there is extensive information on the prevalence of somatic rearrangements and copy number changes in human cancer genomes (from studies using cytogenetics and comparative genomic hybridization) there has previously been limited insight into the prevalence of somatic point mutations[5,6,8–10]. The results of the current studies show that the number of somatic point mutations varies widely both within and between classes of cancer (Fig. 1 and Supplementary Fig. 1).

Seventy-three out of the two-hundred-and-ten cancers showed no somatic mutations at all, whereas others showed exceptionally large numbers (Fig. 1 and Supplementary Fig. 1). The highest mutation prevalence (~77 mutations per Mb) was in two gliomas that were recurrences after treatment with the anticancer drug temozolomide, an alkylating agent that is a known mutagen[7,11,12]. Some individual melanomas and lung cancers also showed substantial numbers of mutations that may relate to the extent of past exposure to ultraviolet radiation (UV) and tobacco smoke carcinogens, respectively. Abnormalities in DNA repair also influenced the number of somatic mutations. Five cancers with defective DNA mismatch repair leading to microsatellite instability had a high prevalence of both base substitutions (14–40 per Mb) and small insertions and deletions at polynucleotide tracts (5–12 per Mb). Occasional cancers without known prior treatment, defects in DNA repair or mutagenic exposure also showed very large numbers of mutations.

Excluding individual cancers with known DNA repair defects or previous treatment, there were differences in overall mutation prevalence between different cancer types (Table 1). Among primary cancers, lung carcinomas sho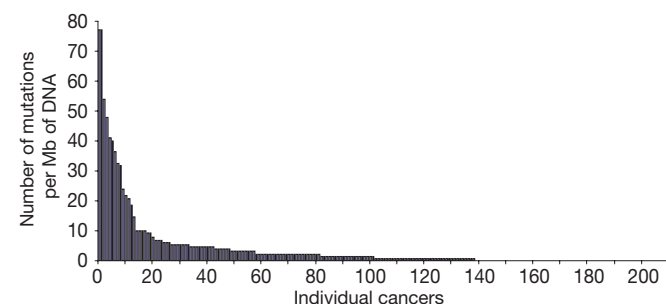wed the highest prevalence of somatic mutations (4.21 per Mb), followed by gastric cancers (2.10 per Mb), ovarian cancers (1.85 per Mb), colorectal cancers (1.21 per Mb, a prevalence similar to that previously reported[10]) and renal cancers (0.74 per Mb). Conversely, testis cancers (0.12 per Mb), lung carcinoids (0 per Mb) and most breast cancers (0.19 per Mb) manifested a much lower prevalence of mutations. The cancer types with high mutation prevalence mainly originate from high turnover, surface epithelia that are subject to recurrent exogenous mutagen exposure (for example, colorectal, lung and gastric). However, other less well understood factors may have a role. For example, the prevalence of somatic mutations in ovarian cancer was higher than that of colorectal cancer. Most ovarian cancers are thought to arise from the specialized peritoneal lining overlying the ovary (or ovarian inclusion cysts deriving from it), for which major exogenous exposures are not recognized and, unlike normal colorectal epithelium, is not thought to be rapidly turning over.

Signatures of somatic mutation

The large numbers of somatic mutations found in this screen also allow comparison of the mutational signatures of cancers. These signatures can carry the specific imprint of previous mutagenic exposures or DNA repair defects and hence provide insights into cancer aetiology. Signatures derived in the past from driver mutations in known cancer genes, notably *TP53* (see http://www-p53.iarc.fr/index.html), have been informative but are inevitably influenced by biological selection, which distorts the patterns generated by the underlying mutational processes. In contrast, in systematic mutation screens most somatic mutations turn out to be passengers (see below) and are therefore not affected by selection.

Mutational signatures differed between cancer types (Fig. 2). In the lung cancers, melanomas and glioblastomas studied they may reflect previous exposure to tobacco carcinogens, UV light and mutagenic alkylating chemotherapy, respectively[6,7]. However, the pathogenesis of other mutational signatures is not understood. For example, we previously showed that a subset of breast cancers has an unusual mutational signature characterized by a high prevalence of C:G>G:C transversions (Fig. 2) that occur in a specific sequence context, at TpC/GpA dinucleotides[5]. We now demonstrate that C:G>G:C changes in lung, ovarian and other cancers are also strongly enriched at TpC/GpA dinucleotides (Table 2), indicating that the underlying mutational process may be more widespread than previously appreciated. In contrast, the TpC/GpA sequence context was not observed in germline C:G>G:C polymorphisms in the protein kinases, suggesting that the process is restricted to cancer cells (Supplementary Table 4). The biological basis of this mutational signature remains

Table 1 | Somatic mutation prevalence by cancer type.

Cancer type	Mutations per Mb of DNA	Number of samples	Number of mutations
ALL	0.57	8	2
Breast	2.70 (†0.19)	16	56
Colorectal	1.21	28	44
Gastric	2.10	18	49
Glioma	22.37 (‡0.32)	9	69
Lung carcinoma	4.21	20	109
Lung carcinoid	0.00	6	0
Ovarian	1.85	25	60
Renal	0.74	23	22
Testis	0.12	13	2
MMR-deficient	32.29	5	209
Melanoma*	18.54	6	144
Other cell lines*	5.64	33	241
All tissues	3.93	210	1,007

ALL, acute lymphoblastic leukaemia; MMR-deficient, mismatch-repair-deficient cancers (two colorectal, two gastric and one ovarian).
* All samples except those indicated are primary cancers or early cultures.
† Removing the single breast cancer PD0119 decreases the breast mutation prevalence to 0.19 per Mb.
‡ Removing temozolomide-exposed PD1487 and PD1489 reduces the glioma mutation prevalence to 0.32 per Mb.
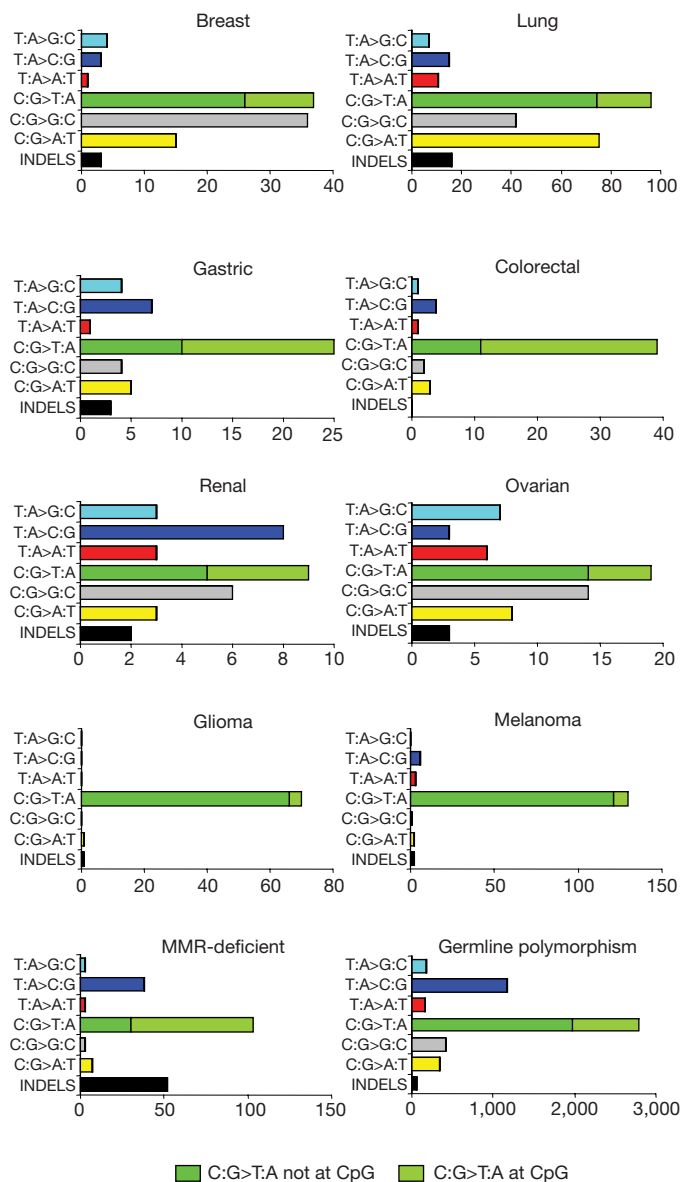

Figure 1 | The prevalence of somatic mutations in human cancer genomes. The number of somatic mutations (base substitutions, insertions/deletions and complex mutations) per Mb of DNA in 210 individual human cancers.

Figure 2 | Mutation spectra of human cancers by tumour type. The numbers of each of the six classes of base substitution and insertion/deletions are shown. C:G>T:A substitutions have been divided into those at CpG dinucleotides and those not at CpG dinucleotides. The data for germline polymorphisms were generated from the protein kinase screen. The data from the two colorectal, two gastric and ovarian cancers that were mismatch-repair-deficient have been shown separately (MMR-deficient).

unknown and may be due to a defect in DNA repair or a shared mutagenic exposure.

Prevalence of driver and passenger mutations

Sequencing the coding exons of the 518 kinases yielded 921 base substitution somatic mutations. These were annotated as non-synonymous (changing an amino acid) or synonymous (not changing

Table 2 | Sequence context of C:G>G:C mutations.

5′ base	Breast	Lung	Others	Germ line	Expected
A	1	6	9	90	20%
C	0	4	4	102	28%
G	0	3	4	114	25%
T	35	29	16	99	26%

Base counts immediately 5′ to cytosine at C:G>G:C somatic mutations and germline variants. The expected percentages were derived from all screened C:G base pairs in the coding sequences of the protein kinases.
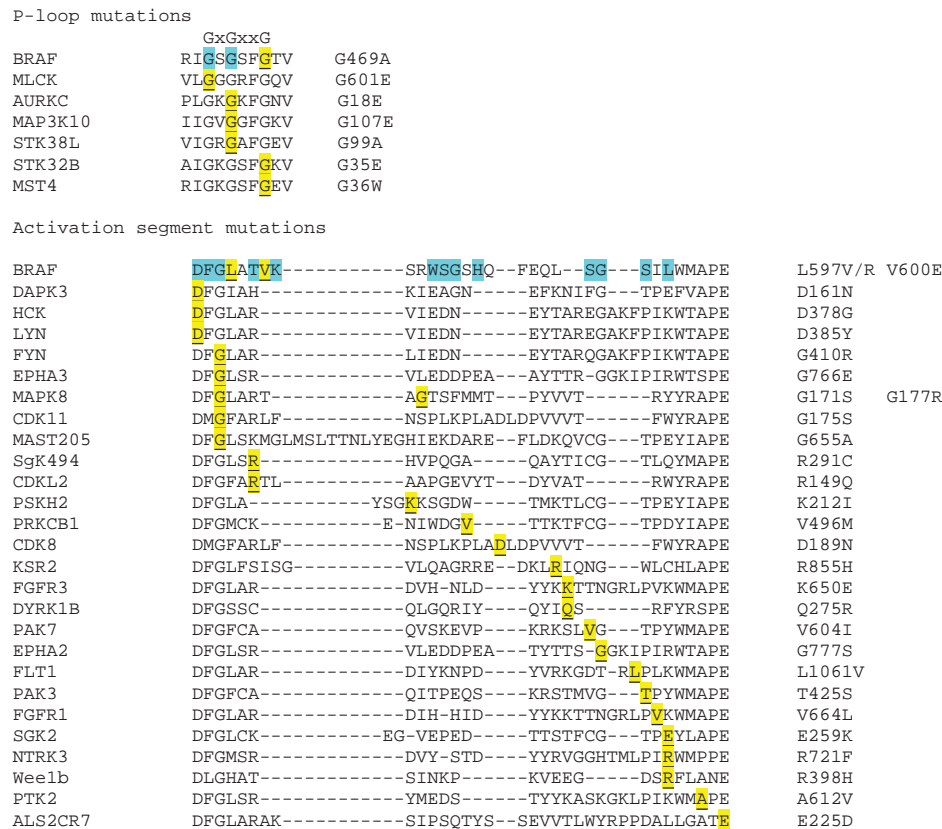
an amino acid). To investigate the numbers of driver and passenger mutations we examined the observed ratio of non-synonymous:synonymous mutations compared with that expected by chance alone[13,14] (see Supplementary Methods for details). The underlying assumption of the analysis is that biological selection is exerted mainly on non-synonymous mutations because these may alter the structure and function of proteins. Conversely, synonymous mutations are generally biologically silent and hence cannot be selected. Therefore, a higher ratio of non-synonymous:synonymous mutations compared with that expected by chance indicates positive selection overall (selection pressure > 1) and is indicative of the presence of driver mutations. A lower non-synonymous:synonymous ratio compared with that expected by chance indicates negative selection overall (selection pressure < 1). This approach has been widely used in studies of selection during evolution[15]. In these analyses we have corrected for several other factors that might influence the non-synonymous:synonymous ratio (see Methods). We are, therefore, interpreting deviation from the expected ratio as owing to selection. However, we cannot completely exclude the existence of other, currently cryptic, factors that might influence the non-synonymous: synonymous ratio and hence imitate the effects of selection.

The selection pressure of all 921 base substitution mutations was 1.29 (95% confidence interval, 1.10–1.51; $P = 0.0013$), demonstrating an excess of non-synonymous mutations compared with that expected and thus providing evidence for the existence of driver mutations within the set. Eleven out of the nine-hundred-and-twenty-one mutations (eight in *BRAF* and three in *STK11*) would have been clearly implicated, on the basis of prior knowledge, in the development of the cancers analysed[16,17]. Removing these mutations, however, only marginally reduces the selection pressure to 1.28 ($P = 0.0025$), indicating that most driver mutations detected were not previously known to be involved in oncogenesis.

To evaluate further the significance of this observation, genes carrying non-synonymous somatic mutations in each cancer type were examined in additional series of each cancer. An additional 454 cancers were examined in this follow-up screen and 91 additional somatic mutations were identified (see Supplementary Information). The selection pressure among this set of mutations was 1.66, indicating that the gene set examined in the follow-up screen was enriched in cancer genes compared with the main screen (selection pressure 1.29, see above), supporting the notion that a proportion of protein kinases harbour oncogenic, driver mutations.

The numbers of passenger and driver mutations present can be estimated from these results (see Supplementary Methods). Of the 921 base substitutions in the primary screen, 763 (95% confidence interval, 675–858) are estimated to be passenger mutations. Therefore, the large majority of mutations found through sequencing cancer genomes are not implicated in cancer development, even when the search has been targeted to the coding regions of a gene family of high candidature. However, there are an estimated 158 driver mutations (95% confidence interval, 63–246), accounting for the observed positive selection pressure. These are estimated to be distributed in 119 genes (95% confidence interval, 52–149). The number of samples containing a driver mutation is estimated to be 66 (95% confidence interval, 36–77). The results, therefore, provide statistical evidence for a large set of mutated protein kinase genes implicated in the development of about one-third of the cancers studied.

Characteristics of driver mutations

To gain further insights into the nature of the driver mutations in protein kinases, we examined how the selection pressure varied among different subsets of mutations. There was no significant difference in selection pressure between mis-sense (1.27), nonsense (1.58) and splice site mutations (1.23) ($P = 0.3363$) or between histological classes of cancer. However, the selection pressure was lower in cancers with defective DNA mismatch repair (MMR) (selection pressure 1.08; $P = 0.72$) compared with MMR-proficient cancers

```
P-loop mutations
                  GxGxxG
BRAF      RIGSGSFGTV      G469A
MLCK      VLGGGRFGQV      G601E
AURKC     PLGKGKFGNV      G18E
MAP3K10   IIGVGGFGKV      G107E
STK38L    VIGRGAFGEV      G99A
STK32B    AIGKGSFGKV      G35E
MST4      RIGKGSFGEV      G36W


Activation segment mutations

BRAF      DFGLATVK----------SRWSGSHQ--FEQL--SG---SILWMAPE    L597V/R V600E
DAPK3     DFGIAH------------KIEAGN-----EFKNIFG---TPEFVAPE    D161N
HCK       DFGLAR------------VIEDN------EYTAREGAKFPIKWTAPE    D378G
LYN       DFGLAR------------VIEDN------EYTAREGAKFPIKWTAPE    D385Y
FYN       DFGLAR------------LIEDN------EYTARQGAKFPIKWTAPE    G410R
EPHA3     DFGLSR------------VLEDDPEA---AYTTR-GGKIPIRWTSPE    G766E
MAPK8     DFGLART-----------AGTSFMMT---PYVVT------RYYRAPE    G171S     G177R
CDK11     DMGFARLF----------NSPLKPLADLDPVVVT------FWYRAPE    G175S
MAST205   DFGLSKMGLMSLTTNLYEGHIEKDARE--FLDKQVCG---TPEYIAPE   G655A
SgK494    DFGLSR------------HVPQGA-----QAYTICG---TLQYMAPE    R291C
CDKL2     DFGFARTL----------AAPGEVYT---DYVAT------RWYRAPE    R149Q
PSKH2     DFGLA----------YSGKKSGDW-----TMKTLCG---TPEYIAPE    K212I
PRKCB1    DFGMCK-----------E-NIWDGV----TTKTFCG---TPDYIAPE    V496M
CDK8      DMGFARLF----------NSPLKPLADLDPVVVT------FWYRAPE    D189N
KSR2      DFGLFSISG---------VLQAGRRE--DKLRIQNG---WLCHLAPE    R855H
FGFR3     DFGLAR------------DVH-NLD----YYKKTTNGRLPVKWMAPE    K650E
DYRK1B    DFGSSC------------QLGQRIY---QYIQS------RFYRSPE     Q275R
PAK7      DFGFCA------------QVSKEVP----KRKSLVG---TPYWMAPE    V604I
EPHA2     DFGLSR------------VLEDDPEA---TYTTS-GGKIPIRWTAPE    G777S
FLT1      DFGLAR------------DIYKNPD----YVRKGDT-RLPLKWMAPE    L1061V
PAK3      DFGFCA------------QITPEQS----KRSTMVG---TPYWMAPE    T425S
FGFR1     DFGLAR------------DIH-HID----YYKKTTNGRLPVKWMAPE    V664L
SGK2      DFGLCK-----------EG-VEPED----TTSTFCG---TPEYLAPE    E259K
NTRK3     DFGMSR------------DVY-STD----YYRVGGHTMLPIRWMPPE    R721F
Wee1b     DLGHAT------------SINKP-----KVEEG-----DSRFLANE     R398H
PTK2      DFGLSR------------YMEDS------TYYKASKGKLPIKWMAPE    A612V
ALS2CR7   DFGLARAK----------SIPSQTYS--SEVVTLWYRPPDALLGATE    E225D
```

Figure 3 | P-loop and activation segment mutations. ClustalW multi-sequence alignment of P-loop and activation segments with all positions of mis-sense mutations highlighted with underline/yellow. Positions of BRAF mutations are shown, with previously identified mutations highlighted in blue and mutations from the current study with underline/yellow. The gene name is indicated on the left. Mutations identified in the study are given to the right of the sequence.

(selection pressure 1.35; $P = 0.00089$). As reported above, MMR-deficient cancers have a higher prevalence of base substitutions than MMR-proficient cancers, presumably due to an increased mutation rate. The lower selection pressure in MMR-deficient cancers is therefore compatible with a model in which driver mutations are overwhelmed by passenger mutations.

Many previously described activating mutations in protein kinase genes that contribute to cancer development are in the kinase domain (see http://www.sanger.ac.uk/genetics/CGP/cosmic/). However, the selection pressure was only slightly higher (1.40) among mutations within kinase domains compared with mutations outside (1.23; $P = 0.08$). Mutations within the P loops and activation segments of kinase domains, in which activating mutations in cancer are often located (Fig. 3), showed a selection pressure of 1.75. Overall, the analysis suggests that, although there may be greater selection pressure for kinase domain mutations, many driver mutations are not in the kinase domains.

There were differences in selection pressure between the ten subclasses[4] of protein kinase ($P = 0.04$) with the highest in calmodulin-dependent protein kinases (1.59), atypical/other kinases (1.32) and tyrosine kinase like kinases (1.33). Many previously reported protein kinase cancer genes have been members of the tyrosine kinase or serine/threonine kinase subclasses. These analyses suggest that other subclasses are also contributing to cancer development.

Potential protein kinase cancer genes

To define further which protein kinases are likely to be carrying driver mutations, the 518 genes have been ranked according to the probability that each is carrying at least one driver mutation, conditional on the selection pressure estimate for each gene (Table 3; Supplementary Table 5; and see Methods). *BRAF* and *STK11* are second and sixteenth in this ranking, providing validation of this

indicator. Remarkably, the gene at the top of this statistical ranking is Titin (*TTN*), which carries 63 non-synonymous and 13 synonymous mutations. The selection pressure associated with *TTN* is only 2.04 compared with 8.36 and 7.16 for *BRAF* and *STK11* respectively and approximately half of the non-synonymous mutations in *TTN* are likely to be passengers. *TTN* is the largest polypeptide encoded by the human genome[18] and has been extensively studied as a component of the muscle contractile machinery. However, it is expressed in many cell types and has other functions that are compatible with a

Table 3 | Protein kinase genes ranked by probability of carrying at least one driver mutation, conditional on the gene-specific selection pressures.

Gene	Ranking (95% confidence interval)	Selection pressure	Number of non-synonymous mutations
TTN	1 (1–3)	2.036	63
BRAF	2 (1–67)	8.362	8
ATM	3 (2–150)	2.920	10
TAF1L	4 (2–145)	3.588	8
ERN1	5 (2–151)	4.538	6
MAP2K4	6 (2–156)	8.665	4
CHUK	7 (2–205)	5.392	5
FGFR2	8 (2–210)	5.096	5
NTRK3	9 (2–518)	4.808	5
MGC42105	10 (2–170)	7.097	4
TGFBR2	11 (2–187)	5.877	4
EPHA6	12 (3–518)	3.949	5
FLJ23074	13 (3–193)	5.403	4
ITK	14 (3–203)	4.887	4
DCAMKL3	15 (3–204)	4.714	4
STK11	16 (3–518)	7.160	3
PAK7	17 (3–518)	4.215	4
STK6	18 (3–518)	6.018	3
BRD2	19 (4–518)	3.773	4
RPS6KA2	20 (4–518)	3.722	4

The top 20 protein kinase genes are shown. See Supplementary Information for the ranking and selection pressures for all 518 genes.

role in oncogenesis[19–21]. The role of *TTN* as a cancer gene is currently a mathematically based prediction and will require direct biological evaluation.

Several genes that are high in the statistical ranking have previously been associated with cancer development. Some of these genes may be activated by their somatic mutations and function as dominant cancer genes, for example *NTRK3* and *ITK*, which are activated by rearrangement in secretory breast cancer and T-cell lymphoma respectively (see http://www.sanger.ac.uk/genetics/CGP/Census/). Others are more likely to be inactivated and operate as recessive cancer genes including *ATM*, in which germline mutations predispose to ataxia telangiectasia[22] and breast cancer[23], *TGFBR2*, in which frameshift somatic mutations are frequently found in mismatch repair deficient cancers[24], and *BMPR1A*, in which germline inactivating mutations cause juvenile polyposis[25]. Each of these three genes has at least one somatic nonsense mutation in the screen. However, most of the genes with probable driver mutations have not previously been associated with cancer development.

Several mutations identified in conserved, functional domains are plausible candidate driver mutations. For example, mutations were found in the glycine residues of the ATP-binding P-loop GxGxxG motif of several protein kinases (Fig. 3). Similar mutations in *BRAF* induce cellular transformation and activate downstream MEK signalling[26]. Mutations were also identified within the activation segment (Fig. 3), a domain frequently harbouring oncogenic mutations in known cancer genes such as *EGFR*, *FLT3*, *KIT* and *BRAF* (see http://www.sanger.ac.uk/genetics/CGP/cosmic/). In particular, the highly conserved DFG motif at the amino-terminal end of the activation segment was mutated in eight protein kinases including three closely related members of the SRC family, *HCK*, *LYN* and *FYN*. Similarly, a Y589H mutation was identified in the juxtamembrane domain of *PDGFRB* in a gastric cancer. *PDGFRB* is activated by translocation in leukaemias (http://www.sanger.ac.uk/genetics/CGP/Census/), and activating mutations in the juxtamembrane domain of the *PDGFRB* paralogue, *PDGFRA*, are found in gastrointestinal stromal tumours (http://www.sanger.ac.uk/genetics/CGP/cosmic/). Tyrosine 589 is highly conserved and mutation of this residue increases the baseline kinase activity of PDGFRB, conferring IL3 independence on BaF3 cells[27].

Clustering of mutations in multiple genes implicates the JNK pathway in cancer development. We and others have identified truncating and mis-sense mutations of *MAP2K4* in lung, colorectal and other cancers[6,28–30]. Downstream signalling from MAP2K4 is mediated, in part, through phosphorylation of MAP2K7 (MKK7) and subsequent activation of JNK1 (MAPK8) and JNK2 (MAPK9)[31,32]. We found two different *MAP2K7* mis-sense mutations of codon 162 (p.R162C and p.R162H) within the kinase domain in colorectal cancers. Moreover, we identified activation segment mutations in *MAPK8* (*JNK1*) and a kinase domain mutation in *MAPK9* (*JNK2*). Taken together, these data indicate that mutations in the JNK pathway are likely to be involved in cancer development.

To investigate formally the distribution of mutated genes with respect to biological pathways, we compared the set of genes with a high probability of having at least one driver mutation to a combined data set of human pathway information that is based on Reactome[33], Panther[34] and INOH[35] data sets. Five-hundred-and-thirty-seven non-redundant pathways containing different combinations of protein kinases were examined. The FGF signalling pathway (Panther Accession P00021 http://www.pantherdb.org/) showed the highest enrichment for kinases containing non-synonymous mutations (corrected *P*-value of 0.011). Among genes in this pathway, previous biological and genetic information suggest that the fibroblast growth factor receptors show several plausible driver mutations. Activating germline mutations of *FGFR3* are known to cause dwarfism[36]. Previous studies have shown that the same amino acids in *FGFR3* that are mutant in the germ line, causing thanatophoric dwarfism, are mutated somatically in bladder cancer[37]. We observed the same

pattern of coincident germline mutations causing skeletal dysplasia and somatic mutations in cancer for *FGFR1* (p.P252T) and *FGFR2* (p.W290C), both in lung cancers[6]. Other mutated genes in the FGF signalling pathway included several MAP kinases such as *MAP2K4*, *MAP2K7*, *MAPK8* (*JNK1*) and *MAPK9* (*JNK2*). Interestingly, pathways involved in apoptosis and cell cycle checkpoints were not enriched in this analysis, although the relative paucity of kinase-domain-containing genes in these pathways limits the power to draw definitive conclusions. Finally, comparison of our results with previously published screens of protein kinases in colorectal cancer[9,30,38] identifies several genes mutated in both colorectal cancer series including *BRAF*, *MAP2K4*, *ERBB4*, *PRKCZ* and *RET*.

Discussion

These large-scale sequencing studies have shown that the prevalence and signature of somatic mutations in human cancers are highly variable. It is likely that the full range of somatic mutation patterns will not be apparent until thousands of cancer samples have been sequenced, each one yielding several dozen mutations each. For some cancers this may require sequencing of hundreds of megabases. This information, however, will ultimately provide major insights into the mutagenic processes underlying neoplastic change.

Our results demonstrate that most somatic mutations in cancer cells are likely to be passenger mutations; however, they have also revealed surprising insights into the number of cancer genes operative in human cancer. Approximately 120 of the 518 genes screened are estimated to carry a driver mutation and therefore function as cancer genes, a larger number than previously anticipated. Interestingly, however, similar conclusions have recently been reached by others. A recent paper reported a mutational analysis of 13,023 genes in 11 colorectal and 11 breast cancers, covering ~1.7 times as much cancer genome as this study[38]. As in this study, they interpret an excess of observed non-synonymous mutations compared with that expected by chance as evidence for the presence of driver mutations. Their design did not include the examination of synonymous changes and hence the analysis of selection pressure undertaken here. Instead, they estimated the expected number of non-synonymous passenger mutations on the basis of prior published data and identified 189 genes that were mutated at significantly higher frequency. Their conclusion was broadly similar, that a large number of cancer-causing mutations and cancer genes are operative in human cancers.

By studying a gene family with a strong track record of involvement in oncogenesis, it is conceivable that we have improved our chances of detecting new cancer genes and that other gene sets may yield a more meagre harvest. Nevertheless, given that we have studied only 518 genes and limited numbers of each cancer type, it seems likely that the repertoire of mutated human cancer genes is larger than previously envisaged. The work presented here suggests that systematic sequencing studies of larger numbers of tumours from a wide variety of cancer types will yield further insights into the development of human cancer, providing new opportunities for molecular diagnosis and therapeutics.

METHODS

DNA was extracted from primary tumours, cancer cell lines and normal tissue samples. Collection and use of tissue samples were approved by the IRB of each institution. Samples estimated to contain more than 80% tumour cells were used. All samples were analysed using Affymetrix 10K SNP arrays to demonstrate that they were from the same individual and to confirm the presence of copy number changes. Microsatellite instability was assessed using the NCI consensus marker panel[39]. PCR primers were designed to amplify all coding exons of the 518 protein kinases[4] annotated in the human genome (available at http://www.sanger.ac.uk/genetics/CGP/). Approximately 10,000 fragments of 500 base pairs were amplified and directly sequenced in both directions from each cancer. Sequence traces were initially evaluated computationally and subsequently manually reviewed. The existence of the variant was then assessed in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) and, if not present, was directly evaluated in normal DNA from the same individual by PCR sequencing using the appropriate

amplimer. Cancer samples showing putative somatic sequence alterations were then re-amplified and re-sequenced along with the appropriate, matched, non-cancer DNA to confirm the somatic nature of the mutation and to eliminate sequencing artefacts. Statistical analyses are outlined in more detail in Supplementary Methods. Deviation of the ratio of non-synonymous:synonymous mutations from that expected by chance was used to indicate the presence of selection on non-synonymous mutations. To assess the significance of this ratio, an exact Monte Carlo test was developed which was applied to the entire set and to subsets of mutations. Additional methods were developed to determine the number of driver mutations, analyse differences in selection between mismatch-repair-deficient and -proficient cancers and to assess the likelihood of a gene being a cancer gene. A combined pathway database was generated by merging Reactome, Panther and INOH to test for the presence of mutated pathways.

1. Futreal, P. A. *et al.* A census of human cancer genes. *Nature Rev. Cancer* **4**, 177–183 (2004).
2. Futreal, P. A. *et al.* Cancer and genomics. *Nature* **409**, 850–852 (2001).
3. Sawyers, C. Targeted cancer therapy. *Nature* **432**, 294–297 (2004).
4. Manning, G., Whyte, D. B., Martinez, R., Hunter, T. & Sudarsanam, S. The protein kinase complement of the human genome. *Science* **298**, 1912–1934 (2002).
5. Stephens, P. *et al.* A screen of the complete protein kinase gene family identifies diverse patterns of somatic mutations in human breast cancer. *Nature Genet.* **37**, 590–592 (2005).
6. Davies, H. *et al.* Somatic mutations of the protein kinase gene family in human lung cancer. *Cancer Res.* **65**, 7591–7595 (2005).
7. Hunter, C. *et al.* A hypermutation phenotype and somatic MSH6 mutations in recurrent human malignant gliomas after alkylator chemotherapy. *Cancer Res.* **66**, 3987–3991 (2006).
8. Bignell, G. *et al.* Sequence analysis of the protein kinase gene family in human testicular germ-cell tumours of adolescents and adults. *Genes Chromosom. Cancer* **45**, 42–46 (2006).
9. Bardelli, A. *et al.* Mutational analysis of the tyrosine kinome in colorectal cancers. *Science* **300**, 949 (2003).
10. Wang, T.-L. *et al.* Prevalence of somatic alterations in the colorectal cancer cell genome. *Proc. Natl Acad. Sci. USA* **99**, 3076–3080 (2002).
11. Lonardi, S., Tosoni, A. & Brandes, A. A. Adjuvant chemotherapy in the treatment of high grade gliomas. *Cancer Treat. Rev.* **31**, 79–89 (2005).
12. Karran, P., Offman, J. & Bignami, M. Human mismatch repair, drug-induced DNA damage, and secondary cancer. *Biochimie* **85**, 1149–1160 (2003).
13. Greenman, C., Wooster, R., Futreal, P. A., Stratton, M. R. & Easton, D. F. Statistical analysis of pathogenicity of somatic mutations in cancer. *Genetics* **173**, 2187–2198 (2006).
14. Yang, Z., Ro, S. & Rannala, B. Likelihood models of somatic mutation and codon substitution in cancer genes. *Genetics* **165**, 695–705 (2003).
15. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol. Biol. Evol.* **11**, 725–736 (1994).
16. Sanchez-Cespedes, M. *et al.* Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
17. Davies, H. *et al.* Mutations of the *BRAF* gene in human cancer. *Nature* **417**, 949–954 (2002).
18. Granzier, H. L. & Labeit, S. Titin and its associated proteins: the third myofilament system of the sarcomere. *Adv. Protein Chem.* **71**, 89–119 (2005).
19. Machado, C. & Andrew, D. J. D-Titin: a giant protein with dual roles in chromosomes and muscles. *J. Cell Biol.* **151**, 639–652 (2000).
20. Machado, C., Sunkel, C. E. & Andrew, D. J. Human autoantibodies reveal Titin as a chromosomal protein. *J. Cell Biol.* **141**, 321–333 (1998).
21. Zastrow, M. S., Flaherty, D. B., Benian, G. M. & Wilson, K. L. Nuclear Titin interacts with A- and B-type lamins *in vitro* and *in vivo. J. Cell Sci.* **119**, 239–249 (2006).
22. Shiloh, Y. ATM and related protein kinases: safeguarding genome integrity. *Nature Rev. Cancer* **3**, 155–168 (2003).
23. Renwick, A. *et al.* ATM mutations that cause ataxia-telangiectasia are breast cancer susceptibility alleles. *Nature Genet.* **38**, 873–875 (2006).
24. Markowitz, S. *et al.* Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. *Science* **268**, 1336–1338 (1995).
25. Howe, J. R. *et al.* Germline mutations of the gene encoding bone morphogenetic protein receptor 1A in juvenile polyposis. *Nature Genet.* **28**, 184–187 (2001).
26. Wan, P. T. C. *et al.* Mechanism of activation of the RAF–ERK signaling pathway by oncogenic mutations of B-RAF. *Cell* **116**, 855–867 (2004).
27. Irusta, P. M. *et al.* Definition of an inhibitory juxtamembrane WW-like domain in the platelet-derived growth factor beta receptor. *J. Biol. Chem.* **277**, 38627–38634 (2002).
28. Teng, D. *et al.* Human mitogen-activated protein kinase kinase 4 as a candidate tumor suppressor. *Cancer Res.* **57**, 4177–4182 (1997).
29. Su, G. *et al.* Alterations in pancreatic, biliary, and breast carcinomas support MKK4 as a genetically targeted tumor suppressor gene. *Cancer Res.* **58**, 2339–2342 (1998).
30. Parsons, D. W. *et al.* Colorectal cancer Mutations in a signalling pathway. *Nature* **436**, 792 (2005).
31. Bogoyevitch, M. A., Boehm, I., Oakley, A., Ketterman, A. J. & Barr, R. K. Targeting the JNK MAPK cascade for inhibition: basic science and therapeutic potential. *Biochim. Biophys. Acta* **1697**, 89–101 (2004).
32. Kyriakis, J. M. & Avruch, J. Mammalian mitogen-activated protein kinase signal transduction pathways activated by stress and inflammation. *Physiol. Rev.* **81**, 807–869 (2001).
33. Joshi-Tope, G. *et al.* Reactome: a knowledgebase of biological pathways. *Nucleic Acids Res.* **33**, D428–D432 (2005).
34. Mi, H. *et al.* The PANTHER database of protein families, subfamilies, functions and pathways. *Nucleic Acids Res.* **33**, D284–D288 (2005).
35. Kushida, T., Takagi, T. & Fukuda, K. Event ontology: a pathway-centric ontology for biological processes. *Pac. Symp. Biocomput.* **11**, 152–163 (2006).
36. Wilkie, A., Patey, S., Kan, S., van den Ouweland, A. & Hamel, B. FGFs, their receptors, and human limb malformations: clinical and molecular correlations. *Am. J. Med. Genet.* **112**, 266–278 (2002).
37. Cappellen, D. *et al.* Frequent activating mutations of FGFR3 in human bladder and cervix carcinomas. *Nature Genet.* **23**, 18–20 (1999).
38. Sjoblom, T. *et al.* The consensus coding sequences of human breast and colorectal cancers. *Science* **314**, 268–274 (2006).
39. Brose, M. S. *et al.* BRAF and RAS mutations in human lung cancer and melanoma. *Cancer Res.* **62**, 6997–7000 (2002).

LETTERS

An ancient nova shell around the dwarf nova Z Camelopardalis

Michael M. Shara[1], Christopher D. Martin[2], Mark Seibert[3], R. Michael Rich[4], Samir Salim[4], David Reitzel[4], David Schiminovich[5], Constantine P. Deliyannis[6], Angela R. Sarrazine[6], Shri R. Kulkarni[2], Eran O. Ofek[2], Noah Brosch[7], Sebastien Lépine[1], David Zurek[1], Orsola De Marco[1] & George Jacoby[8]

Cataclysmic variables (classical novae and dwarf novae) are binary star systems in which a red dwarf transfers hydrogen-rich matter, by way of an accretion disk, to its white dwarf companion[1]. In dwarf novae, an instability[2] is believed to episodically dump much of the accretion disk onto the white dwarf. The liberation of gravitational potential energy then brightens these systems by up to 100-fold every few weeks or months[2]. Thermonuclear-powered eruptions thousands of times more luminous[3,4] occur in classical novae[5], accompanied by significant mass ejection[6] and formation of clearly visible shells[7,8] from the ejected material. Theory predicts that the white dwarfs in all dwarf novae must eventually accrete enough mass to undergo classical nova eruptions[9]. Here we report a shell, an order of magnitude more extended than those detected around many classical novae, surrounding the prototypical dwarf nova Z Camelopardalis. The derived shell mass matches that of classical novae, and is inconsistent with the mass expected from a dwarf nova wind or a planetary nebula. The shell observationally links the prototypical dwarf nova Z Camelopardalis with an ancient nova eruption and the classical nova process.

Z Camelopardalis (Z Cam) is one of the brightest dwarf novae in the sky. At a distance of 163 parsecs (ref. 10; 1 pc = 3.26 light yr = 3.086×10^{16} m), it is also one of the 100 closest of the millions of dwarf novae in the Milky Way. About every 20 days, it brightens by a factor of up to 40 (to apparent visual magnitude ~10), returning to minimum a few days later[11]. The luminosity of Z Cam occasionally hovers in a narrow range between maximum and minimum brightness for weeks to years—it is the prototypical dwarf nova with this behaviour.

The existence of a Z Cam circumstellar shell was suggested[12] on the basis of stationary hydrogen emission line features in the object's visible spectrum. Ultraviolet (UV) spectroscopy[13] and X-ray absorption observations[14] demonstrate that during eruption Z Cam emits a wind. However, no circumstellar matter has ever been imaged in the vicinity of Z Cam.

In the course of its Nearby Galaxy Survey, the NASA orbiting UV telescope GALEX[15] imaged Z Cam and its surroundings. Figure 1 shows the combined UV images, revealing, for the first time, an arc of UV-emitting material with a radius $r \approx 15$ arcmin, centred roughly on, and located southwest (SW) of, Z Cam. Linear nebulosities to the northeast (NE) and southeast (SE) are also seen, covering an area larger than the full Moon. All of these features are detected in multiple GALEX images and pointings. If these nebulosities are associated with Z Cam, they are the largest physical size ($r \approx 0.7$ pc) and largest angular size gas shell ever detected in the environs

of a cataclysmic binary. The next largest ejected shell is that of the classical nova GK Perseii[16] (nova AD 1901) with a radius of 0.1 pc (though it may be surrounded by a much larger fossil planetary nebula). The simplest interpretation of the Z Cam nebulosities (assuming that they arise from Z Cam) is that (1) they are the ejecta
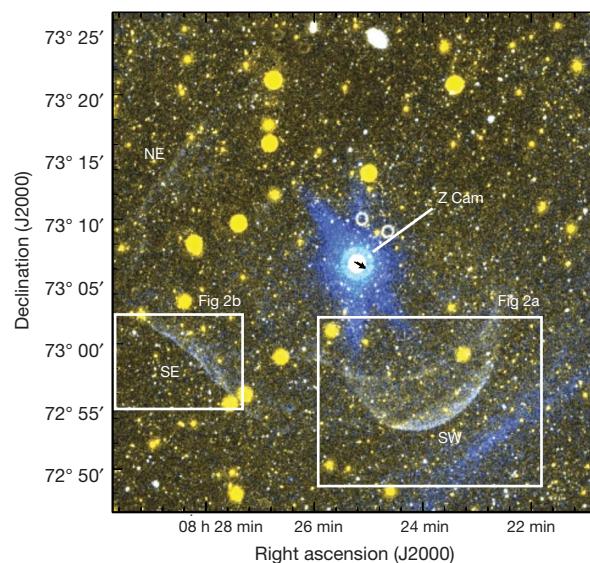
Figure 1 | GALEX satellite image of the field of the dwarf nova Z Camelopardalis on 25 January 2004. Z Cam is the bright star at the centre of the field of view. North is up and east is left. The arcs to the SW and linear features SE and NE of Z Cam are due to shock-heated plasma (see text), while the SW linear feature (more distant from Z Cam than the arcs) is probably due to interstellar cirrus. Two circular features ('doughnuts') seen north and NW of Z Cam are instrumental artefacts, as is the 'aura' surrounding the star out to a radius of 5 arcmin. The GALEX image data include far-UV (effective wavelength $\lambda_{eff} = 151.6$ nm; full-width at half-maximum, FWHM = 25.6 nm) and near-UV ($\lambda_{eff} = 226.7$ nm; FWHM = 73.0 nm) images in circular fields of diameter 1°. The spatial resolution is ~5 arcsec. The imaging data have been processed under the standard GALEX survey pipeline. (Details can be found at http://www.galex.caltech.edu/DATA/gr1_docs). The data in this figure consist of 2,712 s of exposure time combined from two observations of 1,290 and 1,422 s taken on consecutive orbits. The data in both FUV and NUV passband images have been binned 4 × 4 pixels to enhance the low surface brightness features. This figure is a false colour image where red, blue and green colours correspond to the near-UV, far-UV and average of both bands, respectively. The scaling is logarithmic.

[1]Department of Astrophysics, American Museum of Natural History, 79th Street and Central Park West, New York, New York 10024-5192, USA. [2]Department of Physics, Math and Astronomy, California Institute of Technology, 1200 East California Boulevard, Mail Code 405-47, Pasadena, California 91125, USA. [3]Observatories of the Carnegie Institution of Washington, 813 Santa Barbara Street, Pasadena, California 91101, USA. [4]Department of Physics and Astronomy, 430 Portola Plaza, UCLA, Los Angeles, California 90095-1547, USA. [5]Department of Astronomy, Columbia University, New York, New York 10027, USA. [6]Astronomy Department, Indiana University, Bloomington, Indiana 47405-7105, USA. [7]Wise Observatory, Department of Astronomy & Astrophysics, Tel-Aviv University, Ramat-Aviv 69978, Israel. [8]WIYN Observatory, PO Box 26732, Tucson, Arizona 85726, USA.
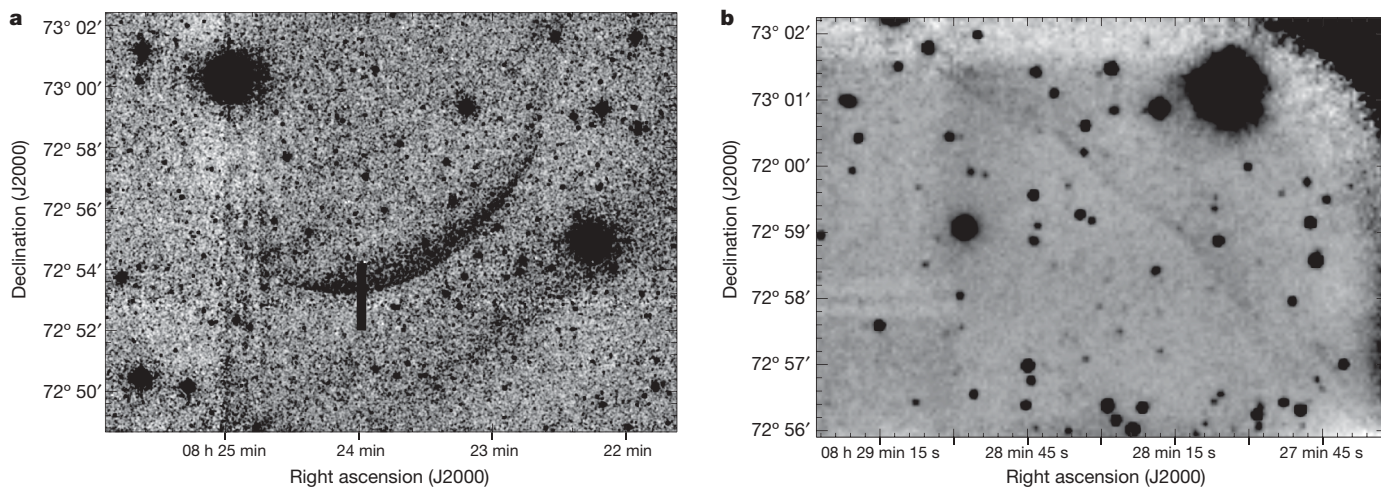
a

b

Figure 2 | Features around Z Cam. a, An image of the arc of emission SW of Z Cam. The location of the long slit used to obtain the spectrum of Fig. 4 is shown superposed on this image. This figure is the sum of 21 images, each of 12 min duration, taken with the WIYN 0.9 m telescope at Kitt Peak National Observatory, Kitt Peak, Arizona. Each image was taken through a narrowband filter with central wavelength 656.2 nm (Hα + [N II]) and FWHM 4.7 nm, on 2 May 2004. The images were flat-fielded and de-biased using the Mosaic data handling software (mscred) in the software package IRAF (Image Reduction and Analysis Facility). Stars on each of the CCD images were used as position markers to match and shift each of the dithered images to the same alignment. The images were then summed (with rejection of cosmic rays and hot pixels) with the package imcalc in IRAF to maximize the signal-to-noise in the final image. **b**, An image of the linear emission-line feature SE of Z Cam. This image is the average of 128 individual images, each of 5 min duration, taken with the Palomar 1.5 m telescope at Mt Palomar, California. Each image was taken through a narrowband filter with central wavelength 656.3 nm and FWHM 10 nm during May through to July 2006. Individual images were binned 4 × 4 pixels before averaging to minimize the effects of seeing differences and focus changes that occur over the run. The images were averaged with the IRAF task imcombine with a sigma clipping so that the highest and lowest values of the 128 images were rejected to remove hot pixels, cosmic rays and dead pixels.

produced by an historic classical nova eruption, or (2) they result from the wind ejected during Z Cam dwarf nova eruptions, or (3) they are a remnant planetary nebula, ejected by the Z Cam white dwarf. We show below that the nebulosities are linked to Z Cam, and that only the classical nova interpretation is consistent with the observations.

Attempts to use the GALEX grism to identify the source of the UV emission were unsuccessful owing to the extended size and low surface brightness of the nebulosity. Broadband red and blue optical images of Z Cam, sensitive enough to detect all the stars imaged by GALEX, show no trace of the GALEX-detected shells. However, narrowband optical images (Fig. 2a and b) sensitive to emission from hydrogen and nitrogen show features that precisely match the shell morphologies of the SW arc and the SE linear feature. This rules out scattered light from interstellar cirrus as the source of the SW arc and the SE linear feature; they are emission-line gas. The linear feature
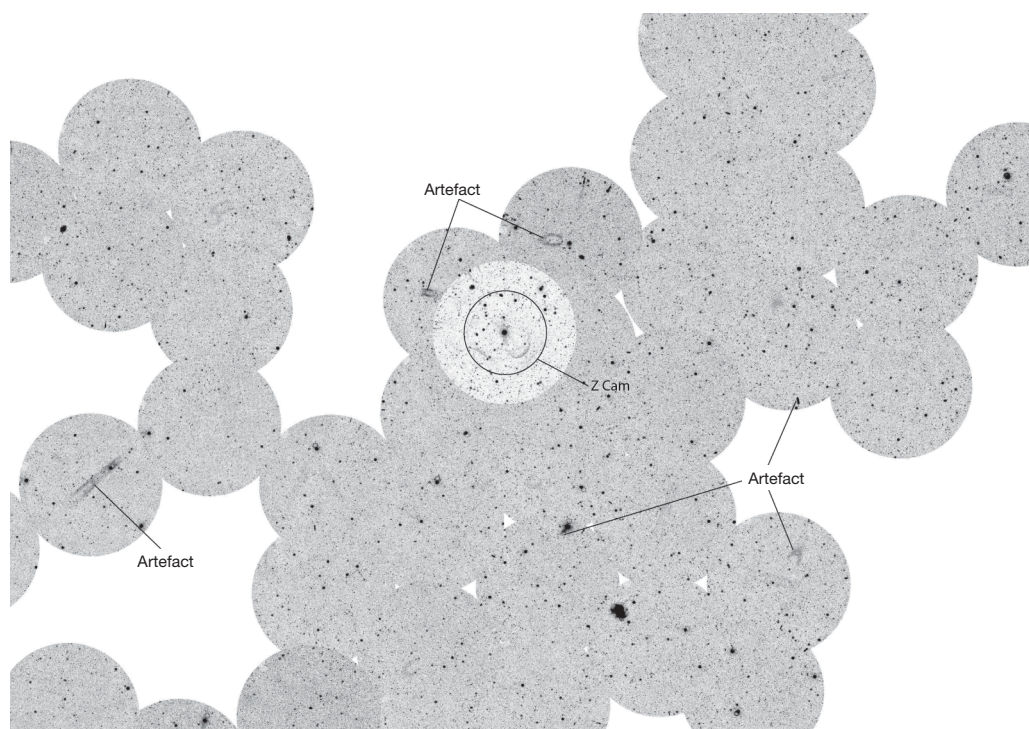
Figure 3 | All fields imaged by GALEX before July 2006, with centres within 200 arcmin of Z Cam. Image is approximately 5.5° × 7.5°. All fields were exposed for ~100 s. The images were processed in the same manner as Fig. 1. Features with the label 'artefact' are instrumental artefacts.

SW of the SW arc is not seen in narrowband images; it is probably interstellar cirrus. Figure 1 demonstrates that the NE linear feature is considerably fainter in the UV than the SE linear feature. Narrowband imagery several times more sensitive than currently available will be needed to detect the NE feature. A mosaic of all available GALEX images in the vicinity of Z Cam (Fig. 3) shows that the SW arc, and NE and SE linear features, are unique (several diffuse 'shells' and elongated features are all instrumental artefacts). More than 16,000 square degrees of sky have been similarly examined (by M. Seibert); the Z Cam UV-emitting features remain unique.

The optical spectrum (Fig. 4) of the SW arc demonstrates the presence of hydrogen, nitrogen and oxygen emission, again ruling out dust reflection by interstellar cirrus as the source of UV light. Gas
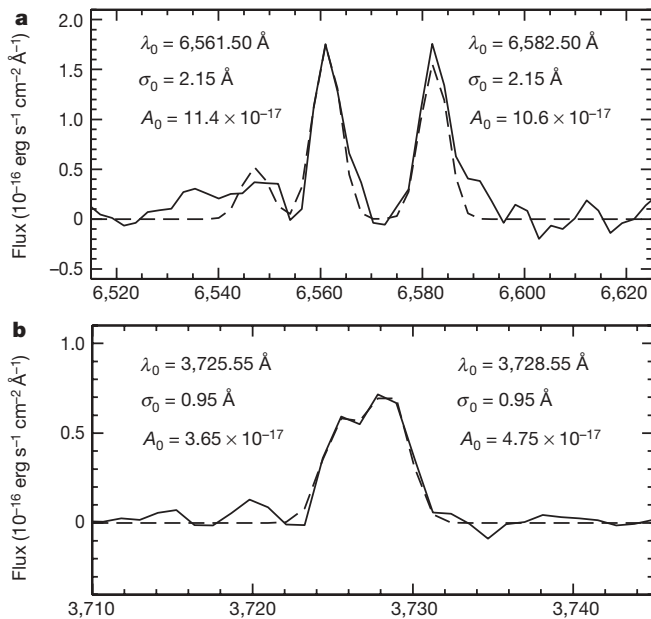
Figure 4 | Emission lines from the Z Cam nebulosity. a, The strongest emission lines in the spectra of the Z Cam nebulosity: forbidden nitrogen [N II] + Hα. Data points are connected by solid lines, and dashed line gaussians are fitted to each line to determine relative intensities. λ_0, σ_0, and A_0 refer, respectively, to the central wavelength, full-width at half-maximum, and flux amplitude of the best fitting gaussians. This is the sum of seven spectra, each of 15 min duration, taken with the Lick 3 m telescope and KAST spectrograph on Mt Hamilton, California, on 7 November 2004. The long-slit spectra imaged a 140″ long (and 2″ wide) slice of sky intersecting the measured position of the nova shell. The resulting spectra were dominated by night-sky emission lines. To determine the location of the nova shell spectrum along the slit, we first calculated a mean spectrum and subtracted it from every position along the slit. An excess in the residuals at the wavelength of the Hα line was found over a section 36″ long, running along the spatial length of the slit, and 18″ off the slit centre; this was due to the emission from the nova shell. A mean spectrum was extracted for that location along the slit. The mean sky spectrum was recalculated, this time excluding the area where nova shell emission was detected. The new mean sky spectrum was subtracted from the extracted spectrum, resulting in a clean, sky-subtracted, nova shell spectrum. Individual emission lines were fitted with gaussian profiles using a least-squares fit. The radial velocity for the nebula was calculated by measuring the centroid of the Hα and [N II] emission lines. The wavelength scale was calibrated locally by using a set of background sky emission lines (air glow) observed between 6,530 Å and 6,635 Å on the same spectrum; the rest wavelengths of the sky lines were obtained from a spectral atlas[25]. **b**, The forbidden oxygen [O II] emission lines in the spectrum of the Z Cam emission. This is the sum of 11 spectra, each of 15 min duration, taken with the National Optical Astronomy Observatory's Mayall 4 m telescope at Kitt Peak National Observatory, Kitt Peak, Arizona, on 8 February 2005. A 300 line mm⁻¹ grating was used with the RC spectrograph. Extraction and calibration of the [O II] emission lines was carried out in similar fashion to the analysis for **a**. The x axes show wavelength in Å.

ejected from Z Cam that moves tangentially to our line of sight should display a radial velocity intermediate between that of Z Cam[12,13] (-38 ± 11 km s⁻¹) and that of matter rotating differentially in the Milky Way at the distance and Galactic latitude ($+32°$) of the star (-3 km s⁻¹). Our spectrum yields heliocentric radial velocities of $V_{rad} = -4 \pm 20$ km s⁻¹ for the hydrogen line, and $V_{rad} = -5 \pm 20$ km s⁻¹ for the strongest nitrogen line. This agreement (within the observational errors) of radial velocities is consistent with the connection between Z Cam and the nebulosities surrounding it.

The lack of an obvious photo-ionization source (Z Cam produces of order 10^{33} erg s⁻¹; it is far too distant and under-luminous) and the line-strengths of forbidden nitrogen lines ([N II]/Hα = 1.25) suggest that the shell is shock-ionized[17]. The [O II] emission line ratio 3,726 Å/3,729 Å ≈ 0.8 in a shocked medium yields[18] a shell density $\rho_{shell} < 10$ cm⁻³, whereas $\rho_{shell} \approx 4$ cm⁻³ is expected in adiabatically shock-compressed interstellar matter (ISM) with a pre-shock density of $\rho_{ISM} = 1$ cm⁻³; we adopt these density values. Figure 2 demonstrates that limb brightening is important in the SW arc, and that the thickness of the shell, dr, is much less than its radius, r. We adopt d$r \approx (10^{-2}–10^{-3})r$ as the shell thickness.

If the shell were spherically symmetric, then its mass would be $M_{shell} \approx 4\pi r^2 \rho_{shell} dr \approx 7 \times (10^{-3}–10^{-4}) M_{Sun}$. The shell's strong asymmetry implies a considerably smaller mass, $M_{shell} \approx (10^{-3}–10^{-4}) M_{Sun}$. Most of this matter must have been 'snow-ploughed' to create the gap between Z Cam and its shell seen in Fig. 1, so the mass ejected from Z Cam itself was significantly less than $(10^{-3}–10^{-4}) M_{Sun}$. The mass of the white dwarf in Z Cam[19] is $(0.99 \pm 0.15) M_{Sun}$. Nova models[5] predict an ejected envelope $M_{env} \approx (3–5) \times 10^{-5} M_{Sun}$ for a $1 M_{Sun}$ white dwarf accreting at the rate deduced for Z Cam[19]. M_{env} could have easily snow-ploughed M_{shell}, as the present-day cavity originally contained up to $M_{ISM} \approx 4\pi r^3 \rho_{ISM} \approx 10^{-2} M_{Sun}$.

The observed range of velocities of matter recently ejected by classical novae is $V_{nova} \approx 300–3,000$ km s⁻¹ (ref. 5). The Z Cam physical shell size then sets an age range of $\tau_{nova} \approx r V_{nova}^{-1} \approx 2,400–240$ yr, with the larger value favoured as significant snow-ploughing has almost certainly taken place. Records of erupting stars two or more millennia ago are almost non-existent, so it is not surprising that no historical record of a Z Cam nova eruption exists.

The rate of steady mass loss from the Z Cam system is deduced from spectrophotometry to be $\dot{M} \approx 10^{-9}–10^{-11} M_{Sun}$ yr⁻¹ (refs 12, 19). If this wind is the source of the shell, then the shell mass deduced above implies a formation age $\tau_{wind} \approx M_{ejecta}/\dot{M} \approx 10^5–10^8$ yr. The time τ_{cool} required for a $1 M_{Sun}$ white dwarf to cool sufficiently such that its luminosity equals that of Z Cam also exceeds 10^6 yr (ref. 20). The proper motion of Z Cam is 21.5 ± 2 mas yr⁻¹ (refs 10, 21), which would have displaced the star by $0.66°–660°$ from its original position during τ_{wind} and by $>6.6°$ during τ_{cool}. The location of Z Cam within a few arcmin of the geometric centre of the SW arc and symmetrically placed with respect to the linear features SE and SW of the star thus argue against both the wind and planetary nebula interpretations.

Irregularities in the interstellar medium surrounding Z Cam before a nova eruption would inhibit the formation of a perfectly symmetric nebula. They would also militate against Z Cam being at the precise geometric centre of arcs or linear features. The Vela supernova remnant[22] displays arcs, linear features and voids with different geometric centres, and reminiscent of those surrounding Z Cam, for just this reason.

The accretion of a sufficiently massive shell onto a white dwarf in any cataclysmic binary, including all dwarf novae, is predicted to inevitably generate pressures large enough to initiate a thermonuclear runaway and hence a classical nova[23] and its shell. This prediction has never been observationally confirmed. The prediction that mass transfer rates in cataclysmic variables should vary on timescales of centuries or millennia, transforming old novae into dwarf novae[9], has but a few exemplars, such as GK Persei[24]. The existence of the extended shell surrounding Z Cam strongly supports both

predictions. It strengthens the claim that cataclysmic variables can undergo cyclic, metamorphic transitions from one sub-type to another. It argues that the present-day, prototypical dwarf nova Z Cam underwent a classical nova eruption a few thousand years ago. During that eruption it must have become, for a few days or weeks, one of the brightest stars in the sky.

1. Warner, B. *Cataclysmic Variable Stars* 7–11 269 (Cambridge Univ. Press, Cambridge, UK, 1995).
2. Osaki, Y. An accretion model for the outbursts of U Geminorum stars. *Publ. Astron. Soc. Jpn* **26**, 429–436 (1974).
3. Lundmark, K. Some facts and suggestions concerning novæ. *Publ. Astron. Soc. Pacif.* **35**, 95–117 (1923).
4. Hubble, E. P. A spiral nebula as a stellar system, Messier 31. *Astrophys. J.* **69**, 103–158 (1929).
5. Yaron, O., Prialnik, D., Shara, M. & Kovetz, A. An extended grid of nova models. II. The parameter space of nova outbursts. *Astrophys. J.* **623**, 398–410 (2005).
6. Payne-Gaposhkin, C. *The Galactic Novae* 53–70 (Interscience Publishers, New York, 1957).
7. Wade, R. A. in *Physics of Classical Novae* (eds Cassatella, A. & Viotti, R.) 179–187 (Proc. IAU Coll. 122, Springer, Berlin, 1990).
8. Downes, R. A. & Duerbeck, H. W. Optical imaging of nova shells and the maximum magnitude-rate of decline relationship. *Astron. J.* **120**, 2007–2037 (2000).
9. Shara, M., Livio, M., Moffat, A. & Orio, M. Do novae hibernate during most of the millennia between eruptions? Links between dwarf and classical novae, and implications for the space densities and evolution of cataclysmic binaries. *Astrophys. J.* **311**, 163–171 (1986).
10. Thorstensen, J. R. Parallaxes and distance estimates for 14 cataclysmic variable stars. *Astron. J.* **126**, 3017–3029 (2003).
11. Mattei, J. A., Waagen, E. O., Oppenheimer, B. D. & Foster, E. G. *Z Camelopardalis Light Curves 1927–1995* (Am. Assoc. Var. Star Observers Monograph, Cambridge, Massachusetts, 1996).
12. Robinson, E. L. Detection of mass loss from the dwarf nova Z Camelopardalis. *Astrophys. J.* **186**, 347–356 (1973).
13. Szkody, P. & Mateo, M. Z. Camelopardalis: Outburst P Cygni profiles and quiescent continuum. *Astrophys. J.* **301**, 286–296 (1986).
14. Baskill, D. S., Wheatley, P. J. & Osborne, J. P. ASCA X-ray observations of the disc wind in the dwarf nova Z Camelopardalis. *Mon. Not. R. Astron. Soc.* **328**, 71–78 (2001).
15. Martin, D. C. *et al.* The Galaxy Evolution Explorer: A space ultraviolet survey mission. *Astrophys. J.* **619**, L1–L4 (2005).
16. Bode, M. F., O'Brien, T. J. & Simpson, M. Echoes of an explosive past: Solving the mystery of the first superluminal source. *Astrophys. J.* **600**, L63–L66 (2004).
17. Sabbadin, F. & D'Odorico, S. Spectroscopic observations of the supernova remnant candidates 3C 400.2 and S91. *Astron. Astrophys.* **49**, 443–445 (1976).
18. Dopita, M. Optical emission from shock waves. *Astrophys. J. Suppl. Ser.* **33**, 437–449 (1977).
19. Knigge, C., Long, K. S., Blair, W. P. & Wade, R. A. Disks, winds, and veiling curtains: Dissecting the ultraviolet spectrum of the dwarf nova Z Camelopardalis in outburst. *Astrophys. J.* **476**, 291–310 (1997).
20. Blocker, T. Stellar evolution of low- and intermediate-mass stars II. Post-AGB evolution. *Astron. Astrophys.* **299**, 755–769 (1995).
21. Høg, E. *et al.* The Tycho-2 catalogue of the 2.5 million brightest stars. *Astron. Astrophys.* **355**, L27–L30 (2000).
22. The Vela supernova remnant. ⟨http://www.aao.gov.au/images/captions/uks002.html⟩ (2002).
23. Shara, M. M. A theoretical explanation of the absolute magnitude-decline time M_B -t_3 relationship for classical novae. *Astrophys. J.* **243**, 926–934 (1981).
24. Simon, V. Dramatic change of the recurrence time and outburst parameters of the intermediate polar GK Persei. *Astron. Astrophys.* **382**, 910–920 (2002).
25. Osterbrock, D. E. *et al.* Night-sky high-resolution spectral atlas of OH and O_2 emission lines for echelle spectrograph wavelength calibration. *Publ. Astron. Soc. Pacif.* **108**, 277–308 (1996).

LETTERS

Meridiani Planum and the global hydrology of Mars

Jeffrey C. Andrews-Hanna[1]*, Roger J. Phillips[1] & Maria T. Zuber[2]

The Opportunity Mars Exploration Rover found evidence for groundwater activity in the Meridiani Planum region of Mars[1,2] in the form of aeolian and fluvial sediments[3] composed of sulphate-rich grains. These sediments appear to have experienced diagenetic modification in the presence of a fluctuating water table[3–5]. In addition to the extensive secondary aqueous alteration, the primary grains themselves probably derive from earlier playa evaporites[1,2,4]. Little is known, however, about the hydrologic processes responsible for this environmental history—particularly how such extensive evaporite deposits formed in the absence of a topographic basin. Here we investigate the origin of these deposits, in the context of the global hydrology of early Mars, using numerical simulations, and demonstrate that Meridiani is one of the few regions of currently exposed ancient crust predicted to have experienced significant groundwater upwelling and evaporation. The global groundwater flow would have been driven primarily by precipitation-induced recharge and evaporative loss, with the formation of the Tharsis volcanic rise possibly playing a role through the burial of aquifers and induced global deformation. These results suggest that the deposits formed as a result of sustained groundwater upwelling and evaporation, rather than ponding within an enclosed basin. The evaporite formation coincided with a transition to more arid conditions[6] that increased the relative impact of a deep-seated, global-scale hydrology on the surface evolution.

Despite its currently cold and desiccated surface, Mars exhibits abundant evidence for a more clement climate in its early history, with surface temperatures above the freezing point of water and liquid precipitation in the mid-latitudes driving runoff and valley formation[7,8]. Evidence from both the global mineralogy[6] and the geomorphology of valley networks[9–11] points towards a transition from conditions in which water was abundant on the surface early in the Noachian epoch (4.55 to ~3.7 Gyr ago[12]), to more arid conditions late in the Noachian and into the Hesperian epoch (~3.7 to ~3.0 Gyr ago[12]). These arid conditions facilitated the formation of stacks of layered sediments rich in evaporitic sulphate minerals and diagenetic haematite at Meridiani Planum and the associated etched terrain[1–5,13]. The contemporaneous formation of the Meridiani deposits with nearby valley networks[14] and the continued formation of sapping valleys[10] suggest that some precipitation persisted even as the climate became more arid, but quickly infiltrated the surface. This precipitation was the dominant contributor to the groundwater hydrology of early Mars, charging aquifers at high elevations[15] (Fig. 1a) in the southern highlands and on the incipient Tharsis volcanic rise.

The Tharsis rise dominates the topography of the western hemisphere of Mars[15], and played a key role in the global geodynamic and hydrologic evolution[16–19]. Although evidence suggests that the Meridiani deposits are ancient, their exact age and timing relative to Tharsis remains unclear. The Noachian age for the deposits suggested by the crater population[5,20,21], coupled with their formation contemporaneous with nearby valley networks[14] (which generally follow a trend consistent with Tharsis-induced deformation[19]), suggests that they may have formed shortly after or towards the end of Tharsis construction. The formation of Tharsis through a combination of extrusive and intrusive volcanism[16–19] largely during the Noachian[19] would have buried any pre-existing aquifers in the region to depths of the order of 30 km beneath the growing volcanic load[19].
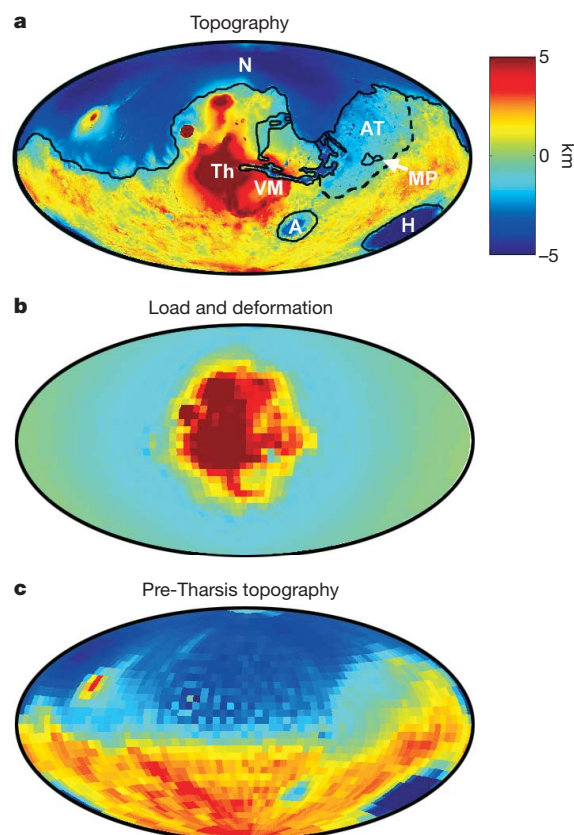

Figure 1 | Topography of Mars and Tharsis-induced deformation. The colour scale applies to all panels. **a,** Topography of Mars from the Mars Orbiter Laser Altimeter, MOLA[15] (Hammer projection centred on 270° E, 0° N). Major features are outlined and labelled, including Tharsis (Th), Valles Marineris canyons (VM), Hellas (H) and Argyre (A) impact basins, the northern lowlands (N), Arabia Terra (AT), and Meridiani Planum (MP). The dashed line represents the break in slope separating Arabia Terra and the southern highlands. **b,** Modelled topography changes due to Tharsis loading and induced deformation (at a resolution of 5° per pixel; assumes a lithosphere thickness of 100 km, Young's modulus of 100 GPa, and Poisson's ratio of 0.25). **c,** Reconstructed pre-Tharsis topography.

[1]McDonnell Center for the Space Sciences and the Department of Earth and Planetary Sciences, Washington University, St Louis, Missouri 63130, USA. [2]Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. *Present address: Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

A limit on the maximum depth to which aquifers are stable is dictated by the brittle–plastic transition (BPT)[22] due to the loss of porosity and permeability in the plastic regime. During the growth of Tharsis, the BPT would have migrated upwards as the deeply buried portions of the crust approached thermal equilibrium. As the buried aquifers underwent the transition to the plastic regime, the water within would have been placed under the burden of the full lithostatic pressure and been expelled, resulting in a steady upwards flux of water beneath Tharsis (Supplementary Fig. 1). This process is analogous to, though on a much larger scale than, the ejection of water from sediments upon burial and compaction in terrestrial environments[23]. If migration of the BPT lagged behind Tharsis growth, as would be expected for plausible thermal diffusion timescales in the deep crust, aquifers beneath Tharsis would have acted as a net hydrological sink during growth of the rise, and then as a net source of fluid after its completion. In addition, the global membrane-flexural deformation of the planet in response to Tharsis loading would have depressed the lithosphere beneath Tharsis and uplifted the antipodal Arabia bulge[19], driving a net migration of groundwater towards Tharsis[14].

We represent global-scale groundwater flow using the time-dependent, nonlinear Boussinesq equation for hydraulic head (h, measured relative to the geoid) in an unconfined aquifer, with a spatially and temporally varying source term (s) representing the flux of water from both surface precipitation and BPT migration beneath Tharsis (Supplementary equation (1)). The large horizontal length scales relative to the aquifer thickness allow us to reduce the problem to two dimensions, using the vertically averaged hydraulic conductivity (K) and storativity (S) over the active aquifer thickness (d), all functions of the elevation and the hydraulic head[22]. The equation is solved in spherical coordinates (θ,φ), using a fully explicit finite difference approach with a grid resolution of 5°. We assume semi-arid conditions, in which the water table is not permitted to rise above the surface, simulating evaporative groundwater loss rather than runoff and the formation of large standing bodies of water. The globally integrated evaporative loss is balanced by precipitation distributed uniformly between latitudes of ±45°, consistent with the observed distribution of valley networks.

We consider three possible scenarios for the global hydrological evolution. The 'pre-Tharsis' model assumes that the Meridiani deposits formed before Tharsis, while the 'post-Tharsis' model assumes that Tharsis was a primordial feature of the planet that pre-dated the deposits. We also model the hydrologic evolution with a steady rate of Tharsis construction over 500 Myr, adjusting the hydraulic head to reflect the changing global topography and geoid. The constructional and membrane-flexural changes in the topography and geoid induced by Tharsis formation are calculated using a spherical harmonic loading model[19,24] and subtracted from the present-day values to reconstruct the intermediate stages of Tharsis growth (Fig. 1). The time-dependent migration of the BPT and ejection of water from deep aquifers (Supplementary Fig. 2) is calculated using a one-dimensional finite difference thermal model.

At the end of the simulation in the pre-Tharsis model, the water table intersects the surface across much of the northern lowlands and in the major impact basins, and these are the main concentrations of evaporative groundwater loss (Fig. 2a). There is also significant groundwater upwelling and evaporation in Meridiani Planum and the surrounding Arabia Terra region, and an isolated location in the southern highlands. Similar results are found for the post-Tharsis model (Fig. 2b), with the exception that the water table is deeper in the Tharsis region as a result of the higher topography there, and there is more extensive groundwater upwelling in the southern highlands. When the growth of Tharsis is included in the model, the lag between Tharsis construction and BPT migration results in a lowering of the water table in the Tharsis region and to a lesser extent in the surrounding southern highlands, preventing groundwater upwelling in these regions (Fig. 2c).

The predicted locations of groundwater upwelling in the northern lowlands and large impact basins would have been largely buried beneath the subsequent sedimentary, volcanic and aeolian material that forms the Hesperian plains. However, outcrops of layered deposits have been identified in the Hellas, Isidis and Argyre basins, and along the dichotomy boundary near Aeolis Mensa[25] (Fig. 2d). In all three models, Meridiani Planum is one of the only regions of currently exposed Noachian-aged crust for which the model predicts
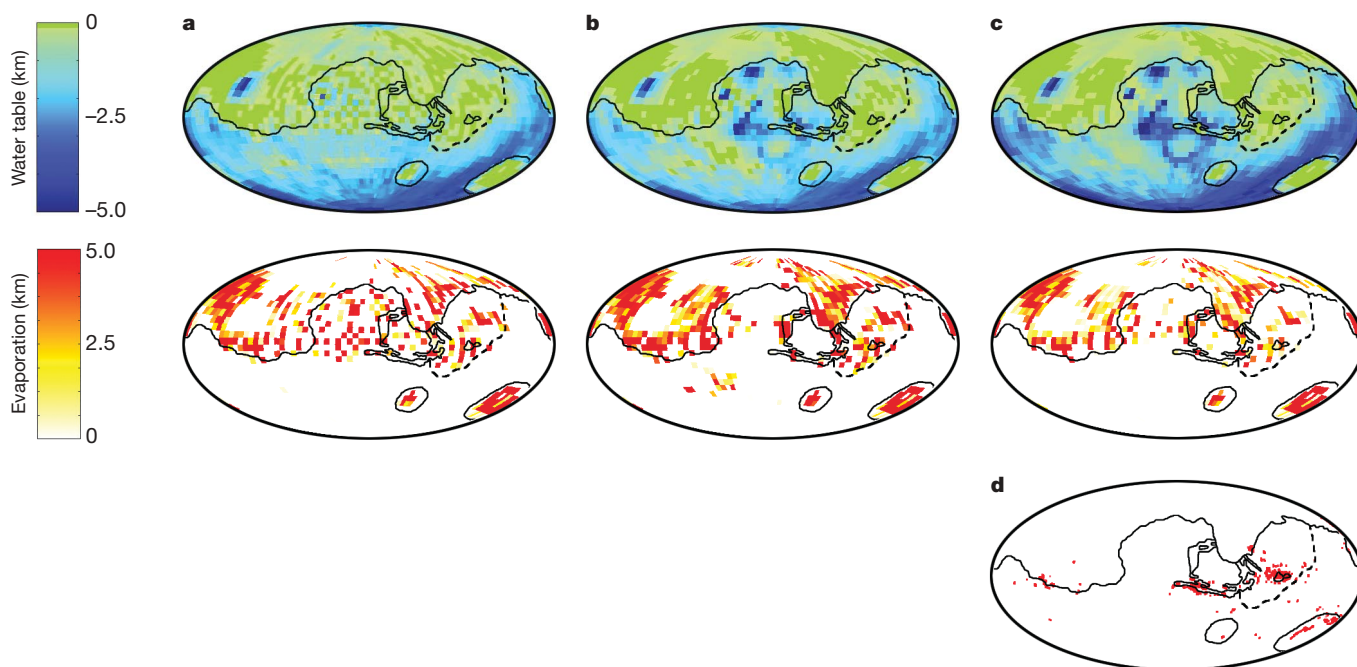

Figure 2 | Modelled hydrologic evolution, and outcrops of layered deposits. a–c, Depth to the water table (top) and cumulative evaporative groundwater loss (bottom); colour scales apply to **a–c.** Results shown after 500 Myr for models assuming pre-Tharsis conditions (**a**), post-Tharsis conditions (**b**) and a steady rate of Tharsis formation (**c**). **d,** The locations of observed outcrops of ancient sedimentary deposits[25] are shown for comparison.

groundwater upwelling and evaporation. The integrated flux to the surface over 500 Myr is equivalent to evaporation of a 2.5-km column of water over a broad area. Increasing the hydraulic conductivity by a factor of 2–5, well within the uncertainty[22], results in similar patterns of upwelling, with 5.4–17 km of evaporation at Meridiani. Assuming that evaporites account for 60% of the deposits[4], their estimated 200–800 m thickness[13,21] would require evaporation of 8–32 km of terrestrial sea water[26]. Fluids deep within the Earth's crust have greater ionic strengths[27] (by a factor of ~3), reducing the amount of water required to ~2.7–11 km, in agreement with the model results.

Although the groundwater fluxes and evaporation rates predicted by the models are small (~10 μm per year), they reflect averaged rates over geologic timescales and large spatial scales. Processes such as local tectonism, volcanic and hydrothermal activity, local-scale precipitation and groundwater flow, climatic fluctuations, variable erosion or deposition rates, or episodic Tharsis formation could modulate this long-term regional groundwater flux. The resulting instantaneous local rates of upwelling could be orders of magnitude greater. The detailed morphology of the deposits would have resulted from the interaction of the regional groundwater upwelling with the local topography and micro-environments at the surface, leading to runoff, ponding, cementation or diagenesis. As the deposits thickened, the water table would rise to remain at or near the surface. Similarly, layered outcrops observed elsewhere[25] may have resulted from groundwater cementation of aeolian or pyroclastic deposits.

To understand the root cause of the groundwater upwelling at Meridiani Planum, we note that it is located near the southern edge of Arabia Terra (Fig. 1). In contrast to the sharp break in slope observed elsewhere along the dichotomy boundary, Arabia Terra is a broad bench perched at intermediate elevation, flanked by steeper slopes leading to the northern lowlands and southern highlands[15,18] (Fig. 3a). For the simplified case of steady, one-dimensional, topography-driven flow in an unconfined aquifer, the discharge is proportional to the product of the height of the water table above the base of the aquifer and the flow velocity, which in turn is proportional to the topographic slope. Conservation of discharge requires that a decrease in slope, and the attendant decrease in flow velocity, will

require an increase in the height of the water table above the base of the aquifer, possibly leading to situations in which the water table attempts to rise above the surface. In the model, this excess groundwater upwelling is balanced by evaporation, returning the water table to the surface. Figure 3b presents a cross-section of the topography, water table elevation, and cumulative evaporative groundwater loss from the model along a great circle passing through Meridiani Planum. Immediately below the first break in slope, at the location of the Meridiani deposits, the water table intersects the surface, resulting in substantial evaporative loss. A similar situation occurs below the second break in slope leading down to the northern lowlands. The intersection of the water table with the surface in Arabia Terra is facilitated by its low topography relative to the highlands, leading to scattered sites of groundwater upwelling and isolated sedimentary deposits throughout this region[21,25,28].

The work presented here sheds light on a previously recognized conundrum regarding the Meridiani evaporites. Terrestrial non-marine evaporites commonly result from the ponding and evaporation of meteoric water within an enclosed or restricted basin, concentrating the solutes leached from rocks over the entire drainage basin within a relatively smaller area[29]. However, no such basin exists in association with the Meridiani deposits[13]. The present results suggest that the Meridiani evaporites formed as a result of sustained groundwater upwelling over a broad region driven by global-scale flow. This groundwater would have leached the solutes from large volumes of aquifer material over a range of depths, before reaching Meridiani and concentrating the solutes at the surface to produce thick deposits of evaporites in the absence of a topographic basin. A small-scale terrestrial analogue is found in the Great Artesian Basin of Australia[30], in which the migration of groundwater over hundreds of kilometres produces local evaporite deposits where the water is released to the surface.

Global mineralogical mapping[6] suggests that Mars experienced a wet period dominated by phyllosilicate formation in the early Noachian, followed by more arid and acidic conditions in the late Noachian and early Hesperian in which the evaporitic sulphate deposits formed. During the earlier wet period, surface runoff and the shallow subsurface hydrology probably masked the signature of the slower, global-scale deep hydrology. The resultant aqueous geochemistry would have been dominated by low-temperature water–rock reactions at the surface, consistent with the observed phyllosilicates. The transition to more arid conditions would have led to the pre-eminence of the global-scale deep groundwater flow over the waning shallow hydrology, with the water chemistry determined by moderate- to high-temperature water–rock reactions at depth, leading to a distinct change in the chemical nature of the surface fluids. Hydrological activity would then be focused in regions of upwelling, with the groundwater interacting with the local surface environment to form the deposits at Meridiani Planum.
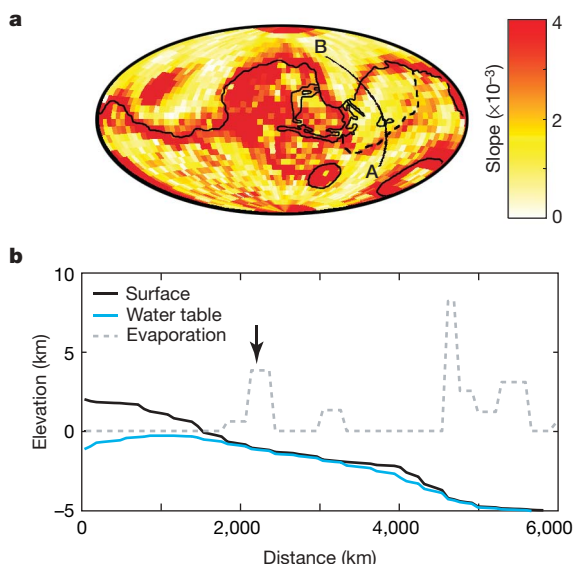
Figure 3 | Role of slopes in driving groundwater upwelling. a, Global slope map of Mars measured at a resolution of 5° per pixel. **b**, Profile of the topography, water table, and the cumulative equivalent column of evaporitic groundwater loss at the end of the Noachian for the steady Tharsis formation model along a great circle path taken parallel to the regional slope in Arabia Terra (line A–B in **a**). The arrow represents the location of Meridiani Planum along the profile. Downturn in the water table at 0 km is a result of the proximity of the Hellas basin.

1. Squyres, S. W. et al. Two years at Meridiani Planum: Results from the Opportunity rover. Science 313, 1403–1407 (2006).
2. Squyres, S. W. et al. Overview of the Opportunity Mars Exploration Rover mission to Meridiani Planum: Eagle Crater to Purgatory Ripple. J. Geophys. Res. 111, E12S12, doi:10.1029/2006JE002771 (2006).
3. Grotzinger, J. P. et al. Stratigraphy and sedimentology of a dry to wet eolian depositional system, Burns formation, Meridiani Planum, Mars. Earth Planet. Sci. Lett. 240, 11–72 (2005).
4. McLennan, S. M. et al. Provenance and diagenesis of the evaporite-bearing Burns formation, Meridiani Planum, Mars. Earth Planet. Sci. Lett. 240, 95–121 (2005).
5. Arvidson, R. E. et al. Nature and origin of the hematite-bearing plains of Terra Meridiani based on analyses of orbital and Mars Exploration Rover data sets. J. Geophys. Res. 111, E12S08, doi:10.1029/2006JE002728 (2006).
6. Bibring, J. P. et al. Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. Science 312, 400–404 (2006).
7. Pollack, J. B., Kasting, J. F., Richardson, S. M. & Poliakoff, K. The case for a wet, warm climate on early Mars. Icarus 71, 203–224 (1987).
8. Hynek, B. M. & Phillips, R. J. New data reveal mature, integrated drainage systems on Mars indicative of past precipitation. Geology 31, 757–760 (2003).

9. Baker, V. R. & Partridge, J. B. Small martian valleys: Pristine and degraded morphology. *J. Geophys. Res.* **84,** 3561–3572 (1986).
10. Williams, R. M. E. & Phillips, R. J. Morphometric measurements of martian valley networks from Mars Orbiter Laser Altimeter (MOLA) data. *J. Geophys. Res.* **106,** 23737–23751 (2001).
11. Harrison, K. P. & Grimm, R. E. Groundwater-controlled valley networks and the decline of runoff on early Mars. *J. Geophys. Res.* **110,** E12S16, doi:10.1029/2005JE002455 (2005).
12. Hartmann, W. K. & Neukum, G. Cratering chronology and the evolution of Mars. *Space Sci. Rev.* **96,** 165–194 (2001).
13. Hynek, B. M., Arvidson, R. E. & Phillips, R. J. Geologic setting and origin of the Terra Meridiani hematite deposit on Mars. *J. Geophys. Res.* **107,** 5088, doi:10.1029/2002JE001891 (2002).
14. Phillips, R. J. & Hynek, B. M. THEMIS and MOLA provide 3-D stratigraphy in the Terra Meridiani region of Mars. *Eos* 86(*Fall Meet.*), abstr. P24A–04 (2005).
15. Smith, D. E. *et al.* Mars Orbiter Laser Altimeter (MOLA): Experiment summary after the first year of global mapping of Mars. *J. Geophys. Res.* **106,** 23689–23722 (2001).
16. Solomon, S. C. & Head, J. W. Evolution of the Tharsis province of Mars—The importance of heterogeneous lithospheric thickness and volcanic construction. *J. Geophys. Res.* **87,** 9755–9774 (1982).
17. Phillips, R. J., Sleep, N. H. & Banerdt, W. B. Permanent uplift in magmatic systems with application to the Tharsis region of Mars. *J. Geophys. Res.* **95,** 5089–5100 (1990).
18. Zuber, M. T. *et al.* Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* **287,** 1788–1793 (2000).
19. Phillips, R. J. *et al.* Ancient geodynamics and global-scale hydrology on Mars. *Science* **291,** 2587–2591 (2001).
20. Lane, M. D., Christensen, P. R. & Hartmann, W. K. Utilization of THEMIS visible and infrared imaging data for crater population studies of the Meridiani Planum landing site. *Geophys. Res. Lett.* **30,** 1770, doi:10.1029/2003GL017183 (2003).
21. Edgett, K. S. The sedimentary rocks of Sinus Meridiani: Five key observations from data acquired by the Mars Global Surveyor and Mars Odyssey orbiters. *Mars* **1,** doi:10.1555/mars.2005.0002 (2005).
22. Hanna, J. C. & Phillips, R. J. Hydrological modeling of the Martian crust with application to the pressurization of aquifers. *J. Geophys. Res.* **110,** E01004, doi:10.1029/2004JE002330 (2005).
23. Deming, D., Nunn, J. A. & Evans, D. G. Thermal effects of compaction-driven flow from overthrust belts. *J. Geophys. Res.* **95,** 6669–6683 (1990).
24. Banerdt, W. B. Support of long-wavelength loads on Venus and implications for internal structure. *J. Geophys. Res.* **91,** 403–419 (1986).
25. Malin, M. P. & Edgett, K. S. Ancient sedimentary rocks of early Mars. *Science* **290,** 1927–1937 (2000).
26. Handford, C. R. in *Evaporites, Petroleum, and Mineral Resources* (ed. Melvin, J. L.) 1–66 (Elsevier, New York, 1991).
27. Moller, P. *et al.* Paleofluids and recent fluids in the upper continental crust: Results from the German Continental Deep Drilling Program (KTB). *J. Geophys. Res.* **102,** 18233–18254 (1997).
28. Hynek, B. M. Implications for hydrologic processes on Mars from extensive bedrock outcrops throughout Terra Meridiani. *Nature* **431,** 156–159 (2004).
29. Smoot, J. P. & Lowenstein, T. K. in *Evaporites, Petroleum, and Mineral Resources* (ed. Melvin, J. L.) 189–347 (Elsevier, New York, 1991).
30. Habermehl, M. A. The Great Artesian Basin, Australia. *J. Austr. Geol. Geophys.* **5,** 9–38 (1980).

Observation of the two-channel Kondo effect

R. M. Potok[1,3]†, I. G. Rau[2], Hadas Shtrikman[4], Yuval Oreg[4] & D. Goldhaber-Gordon[1]

Some of the most intriguing problems in solid-state physics arise when the motion of one electron dramatically affects the motion of surrounding electrons. Traditionally, such highly correlated electron systems have been studied mainly in materials with complex transition metal chemistry[1,2]. Over the past decade, researchers have learned to confine one or a few electrons within a nanometre-scale semiconductor 'artificial atom', and to understand and control this simple system in detail[3]. Here we combine artificial atoms to create a highly correlated electron system within a nano-engineered semiconductor structure[4]. We tune the system *in situ* through a quantum phase transition between two distinct states, each a version of the Kondo state[5], in which a bound electron interacts with surrounding mobile electrons. The boundary between these competing Kondo states is a quantum critical point—namely, the exotic and previously elusive two-channel Kondo state[6,7], in which electrons in two reservoirs are entangled through their interaction with a single localized spin.

The Kondo effect has become a hallmark of coherent electron transport in a variety of nanostructures, ranging from lithographically defined semiconductors to carbon nanotubes and molecules[8,9]. Kondo applied the phenomenological hamiltonian[10]

$$H_K = J \, \mathbf{s} \cdot \mathbf{S} + H_{\text{reservoir}} \qquad (1)$$

to describe a magnetic impurity embedded in a host sea of electrons. A localized spin $\mathbf{S}$ couples antiferromagnetically with strength J to spins $\mathbf{s}$ of electrons in the surrounding reservoir. $H_{\text{reservoir}}$ represents the free electrons in the reservoir. At temperatures below the Kondo temperature T_K, electrons in the reservoir screen the localized spin. The Kondo hamiltonian was later found to be derivable from the more microscopic Anderson model, which consists of an electron bound to an impurity site in a metal host (Fig. 1a). Here, Kondo's antiferromagnetic coupling emerges from tunnelling on and off the local site.

Many systems of strongly interacting particles can be understood in the framework of Landau's Fermi liquid theory, whose basic entities, termed quasiparticles, roll most effects of interactions into changes in particle properties, such as mass and energy. Although the Kondo ground state is complex, its excitations can still be described as weakly interacting quasiparticles. Some of the most intriguing problems in solid-state physics arise when this simplification cannot be applied. Examples of such highly correlated systems include Luttinger liquids, fractional quantum Hall Laughlin liquids, high-temperature superconductors, and the two-channel Kondo system, a novel state studied experimentally in this Letter.

In the two-channel Kondo (2CK) model, introduced 25 years ago in refs 6 and 7 independently, a localized spin $\mathbf{S}$ is antiferromagnetically coupled to two independent reservoirs of electrons according to the hamiltonian:

$$H_{2CK} = J_1 \mathbf{s}_1 \cdot \mathbf{S} + J_2 \mathbf{s}_2 \cdot \mathbf{S} + H_{\text{reservoirs}} \qquad (2)$$

The symmetric 2CK state is formed when the two independent channels (or reservoirs) are equally coupled to the magnetic impurity, that

is, $J_1 = J_2$. Each reservoir individually attempts to screen the local spin. As they cannot both screen the spin, this is an unstable situation, resulting in a new ground state in which the local impurity is only partially screened. Unlike the situation for single-channel Kondo (1CK), in the 2CK state the quasiparticle concept of Fermi liquid theory does not apply: the decay rate for a low energy excitation ($\sim\sqrt{\varepsilon}$) is greater than the energy ε of the excitation itself. Stable low-lying excitations must thus be collective[2,11].

Any difference in channel coupling will force the system away from the non-Fermi-liquid 2CK state and towards the 1CK state associated with the more strongly coupled reservoir. Although the symmetric 2CK state might therefore seem difficult to access, it has been invoked to explain remarkable low-energy properties of some heavy fermion materials[12–14] and glassy metals[15–17]. However, the connections of these experimental observations to 2CK theory remain problematic, in part because the microscopic electronic structure of the various materials is unclear.

In this Letter, we present experimental results on an artificial impurity that is designed to display the 2CK effect. Crucially, we can precisely control the coupling constants J_1 and J_2, while maintaining the independence of the two channels. The system is built from a GaAs/AlGaAs heterostructure containing a low density ($n_e = 2 \times 10^{11}\,\text{e}^-\,\text{cm}^{-2}$), high mobility ($\mu = 2 \times 10^6\,\text{cm}^2\,\text{V}^{-1}\,\text{s}^{-1}$) two-dimensional electron gas (2DEG) 68 nm below the surface. We follow the proposal of ref. 4 to produce two independent screening channels for an artificial magnetic impurity. A gate-defined quantum dot containing ~25 electrons in an area of 0.04 μm^2 plays the role of our magnetic impurity (Fig. 1d, left). Its bare charging energy $U \approx 1\,\text{meV}$ and its average single-particle level spacing $\Delta \approx 100\,\mu\text{eV}$. Previous experiments that claim to probe 2CK behaviour[12–17] used a local orbital degeneracy in place of spin, freeing spin of the surrounding conduction electrons to act as the channel index. In contrast, our local degeneracy is a real spin, and we use two physically separated reservoirs (red and blue in Fig. 1d) for the screening channels. Two leads of the small quantum dot cooperate as a single screening channel with antiferromagnetic coupling J_{ir} ('infinite reservoir')[18]. An additional lead is made finite in size (red, Fig. 1d), so that adding or removing an electron from this reservoir is energetically forbidden at low temperature, a phenomenon known as Coulomb blockade. The area of the finite reservoir is ~3 μm^2, corresponding to a charging energy $E_c = 100\,\mu\text{eV} \approx 1.2\,\text{K}$, and a single-particle level spacing $\Delta_{\text{fr}} = 2\,\mu\text{eV} \approx 25\,\text{mK}$ ('finite reservoir'). This level spacing is only slightly larger than the base electron temperature of our dilution refrigerator (12 mK, as determined by Coulomb blockade thermometry on the small quantum dot), and indeed we cannot resolve these levels even at base temperature. Hence, the finite reservoir has an effectively continuous density of states and can screen the magnetic impurity. As Coulomb blockade prevents exchange of electrons with the other leads, the finite reservoir acts as a second Kondo screening channel (Fig. 1e), with antiferromagnetic coupling J_{fr},

[1]Department of Physics, [2]Department of Applied Physics, Stanford University, Stanford, California 94305, USA. [3]Department of Physics, Harvard University, Cambridge, Massachusetts 02138, USA. [4]Department of Condensed Matter Physics, Weizmann Institute of Science, Rehovot 96100, Israel. †Present address: Advanced Micro Devices, Austin, Texas 78741, USA.

allowing the possibility of observing and studying the 2CK effect. The 2CK hamiltonian (equation (2)) has three possible ground states, depending on the relative couplings to the two reservoirs: 1CK with the finite reservoir ($J_{fr} > J_{ir}$), 1CK with the infinite reservoir ($J_{ir} > J_{fr}$), and 2CK at the quantum critical point ($J_{fr} = J_{ir}$).

In Fig. 2, we demonstrate that the small quantum dot can act as a tunable magnetic impurity and display the 1CK effect. If the small quantum dot has an odd number of electrons, it has a net spin and acts as a magnetic impurity. With gate 'n' de-energized (0 V), the system has three conventional leads (blue and red in Fig. 1d), all of which cooperate to screen the magnetic impurity with a single energy scale kT_K. At temperature $T \lesssim T_K$, the Kondo effect enhances scattering and hence conductance from one lead to another. We measure the conductance $g \equiv dI/dV_{ds}|_{V_{ds}=0}$ between the two blue leads (I is current, and V_{ds} is voltage between source and drain reservoirs). As temperature is increased, the Kondo state is partially destroyed, so the conductance decreases (Fig. 2b). The conductance as a function of temperature (for example, Figure 2b inset) matches the expected form $g(T)$ for a quantum dot in the Kondo regime[19,20]; see Supplementary Information for complete analysis. This, and all other measurements reported here, are performed in a magnetic field $B = 130$ mT normal to the plane of the heterostructure. The orbital
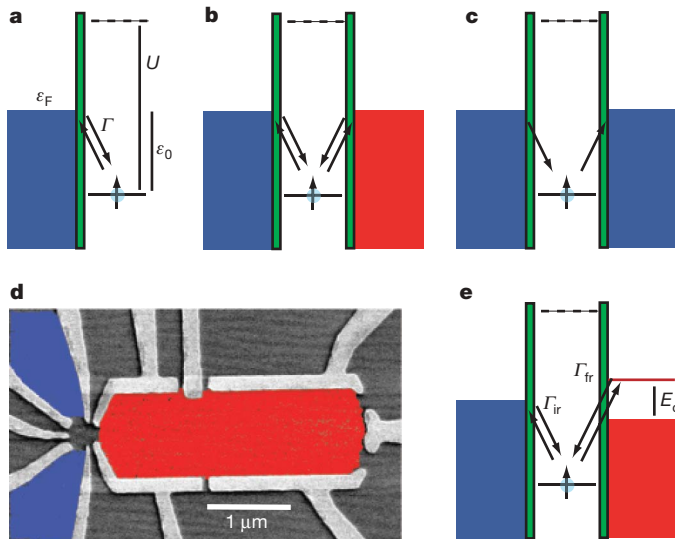
Figure 1 | One and two-channel Kondo effects. a, Single channel Kondo (1CK) effect. The Anderson model describes a magnetic impurity in a metal as a single spin-degenerate state (right side of green barrier) coupled to a Fermi reservoir of electrons (left) with Fermi energy ε_F. Coulomb interaction U between localized electrons favours having only a single electron in the localized state. The antiferromagnetic coupling J between the localized spin and the reservoir depends on the tunnelling rate Γ, the depth of the level ε_0, and U, according to $J \propto \Gamma U/[\varepsilon_0(\varepsilon_0 + U)]$ (ref. 30). At low temperature, high-order tunnelling processes between the local state and the Fermi reservoirs coherently add together to screen the localized electron spin. **b,** Two channel Kondo (2CK) effect. A localized electron is now coupled to two independent Fermi reservoirs (blue and red). If the two independent reservoirs are equally coupled to the localized spin, each will individually attempt to screen the spin, resulting in the formation of a highly correlated electron state. **c,** Physically separating two reservoirs does not suffice to make them independent. If a localized electron can hop off the site to the right reservoir and a new electron can hop onto the site from the left, the two reservoirs will cooperate in screening the localized spin. To create two independent screening channels, processes that transfer electrons from one reservoir to another must be suppressed. **d,** Experimental realization of the 2CK effect. We add an additional finite reservoir (red) to an artificial magnetic impurity connected to an infinite reservoir composed of two conventional leads (blue). **e,** Coulomb blockade suppresses exchange of electrons between the finite reservoir and the normal leads at low temperature. The two reservoirs (blue and red) hence act as two independent screening channels (see main text).

effect of this modest field suppresses direct transmission through the small quantum dot, which we found to yield Fano lineshapes at zero magnetic field (compare ref. 21). Owing to the small g-factor of electrons in GaAs/AlGaAs heterostructures, $|g| \approx 0.4$, the Zeeman effect of the field is unimportant in both 1CK and 2CK regimes (see Supplementary Information for details). All results presented in this Letter are for this same electron occupancy, although we have observed similar behaviour in the next Kondo valley (two fewer electrons in the small dot), as well as on thermally cycling the device.

Figure 3 explores the effect of energizing gate 'n', thus forming the finite reservoir. Differential conductance $g(T, V_{ds}) = dI/dV_{ds}$ is enhanced near zero bias (Fig. 3b and f) when the electrostatic potential of the small dot is set to the middle of the Kondo valleys in Fig. 2b or c, respectively. This is a manifestation of the enhanced density of states at the Fermi level, widely accepted as one of the classic signatures of the Kondo effect, demonstrating clearly that the small dot acts as a magnetic impurity. Remarkably, the zero-bias enhancement changes to zero-bias suppression as gate n is made more negative, closing off the big dot to form a finite reservoir with integer occupancy (Fig. 3g). The change signals that the single-channel Kondo state with the leads has been broken, to form instead solely with the finite reservoir. This occurs for $J_{fr} > J_{ir}$, as shown in more detail in Fig. 3h and Supplementary Information. With slightly weaker coupling to the finite reservoir (Fig. 3c), $J_{ir} > J_{fr}$, the Kondo state is formed solely with the infinite reservoir. This effect requires the finite reservoir to have integer occupancy, that is, the device must be set to a Coulomb blockade valley of the finite reservoir.

In Fig. 3d and h, we provide further evidence that, with the finite reservoir formed, two independent 1CK states can exist, depending on the relative coupling of the small dot to the two reservoirs. We
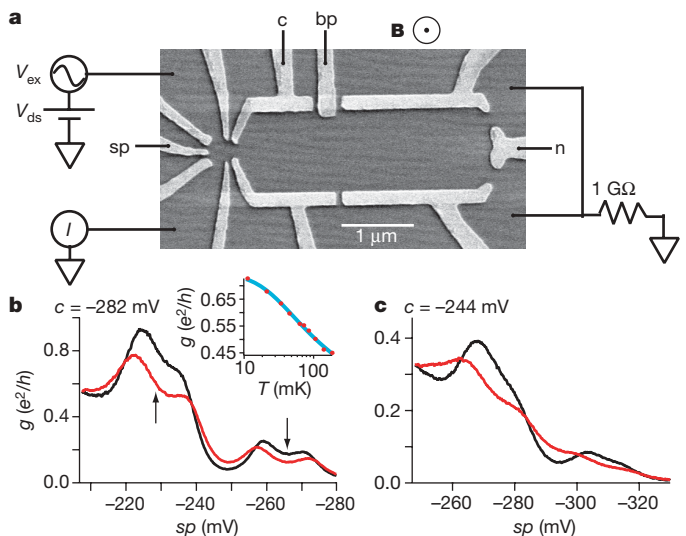

Figure 2 | Artificial magnetic impurity. a, Scanning electron micrograph of a device similar to that measured. The device consists of a small quantum dot (magnetic impurity, left) coupled to conventional leads (top and bottom left) and to a large quantum dot (finite reservoir, right). Electrons are depleted under every gate by application of negative voltages. In the experiments described here, voltages are varied on gates labelled c ('coupling' between dots), bp ('big dot plunger'), sp ('small dot plunger') and n ('nose', which opens or closes the big dot). All transport measurements presented in Figs 3 and 4 are measured with source and drain connected as shown, in a magnetic field of 130 mT normal to the plane of the 2DEG. **b,** With gate voltage $n = 0$, the large dot opens into an infinite reservoir. Arrows mark regions where the small dot has an unpaired spin, leading to enhanced conductance at 12 mK (black) compared to 50 mK (red). Fitting the temperature dependence of the conductance (**b** inset), we find that the Kondo temperature ranges from 110 to 300 mK (see Supplementary Information). In **c,** the data from **b** are shown for stronger tunnel coupling to the right lead: $c = -244$ mV instead of -282 mV. From the temperature dependence of **c,** we find that T_K ranges from 180 to 320 mK.

have fine control of the occupancy of both the finite reservoir and the small dot with gates bp and sp (see Fig. 2 for gates; this fine control is demonstrated in Supplementary Information). The differential conductance $g(V_{ds}, T)$ of a 1CK system is expected to follow a specific form as a function of both bias and temperature, at an energy scale substantially below kT_K (ref. 22):

$$\frac{g(0, T) - g(V_{ds}, T)}{T^{\alpha}} = \kappa \left(\frac{eV_{ds}}{kT}\right)^2 \quad (3)$$

where the exponent $\alpha = 2$ is characteristic of 1CK behaviour, and $\kappa = 0.82 g_0 / T_K^2$. The numerical prefactor of order unity is dependent on the underlying model, numerical calculations, and proximity to
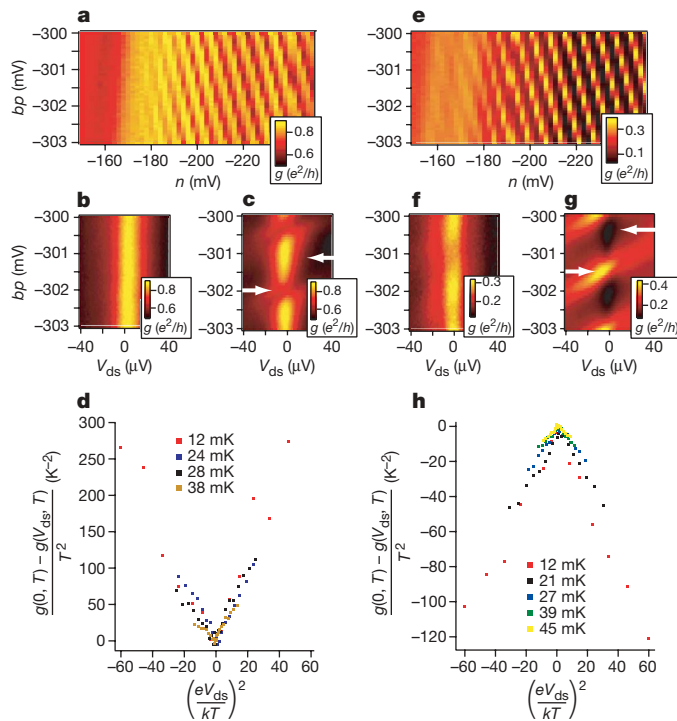
Figure 3 | The formation of two competing 1CK states with two different reservoirs. a, sp is set so that the small dot has an unpaired spin (middle of a Kondo valley), and c is set to −282 mV. Conductance is shown as a function of n and bp. With n > −180 mV, varying bp has little effect on the conductance. As n is made more negative, the reservoir becomes finite and a series of stripes in the conductance map reflects Coulomb blockade in the finite reservoir (see Supplementary Information for explanation of the modulation). In **b** and **c**, differential conductance as a function of bias (V_{ds}) and bp is shown for n = −170 mV and n = −224 mV, respectively. In **b**, there is clear zero-bias enhancement, consistent with the Kondo effect. In **c**, the zero-bias enhancement (denoted by the right-hand arrow) is modulated by the charge state of the finite reservoir (see Supplementary Information). The left-hand arrow marks the charge degeneracy point of the finite reservoir. **d,** Plotting a specific combination of temperature and bias collapses the data for a single point in occupancy of the two dots (small dot with odd occupancy and large dot in Coulomb blockade; T_K = 175 mK, $g_0 = 0.75e^2/h$) onto a single V-shaped curve, corresponding to the scaling relation predicted for 1CK behaviour (equation (3)). **e,** Same as **a**, except with stronger interdot coupling: c = −244 mV. **f,** As in **b**, we observe a clear zero-bias conductance enhancement for n = −170 mV. **g,** However, in this case at n = −224 mV the zero-bias enhancement is replaced by zero-bias suppression (denoted by the right-hand arrow). Here the local spin forms a Kondo state with the finite reservoir, suppressing conductance through the small quantum dot (see main text). Again, the zero-bias feature is modulated by the charge state of the finite reservoir, and the left-hand arrow marks a charge degeneracy point of the finite reservoir. **h,** We collapse differential conductance data at a single point in (sp, bp) onto a single inverted V-shaped curve using the same temperature–bias scaling as in **d**. Deviations from perfect scaling may be related to the slightly lower Kondo temperature (T_K = 120 mK, $g_0 = 0.16e^2/h$.)

the symmetric 2CK fixed point (see Supplementary Information), so we simply treat κ as a free fitting parameter for each set of gate voltages. Figure 3d demonstrates excellent 1CK scaling at temperatures of 12, 24, 28 and 38 mK, all well below T_K. A nonlinear fit to the data in Fig. 3d yields $\alpha = 1.72 \pm 0.40$ (95% confidence limits), consistent with $\alpha = 2$.

In Fig. 3h, we demonstrate that at stronger coupling to the finite reservoir ($J_{fr} > J_{ir}$), the small dot forms a Kondo state with the finite reservoir, as manifested by low-energy suppression rather than enhancement of conductance between the normal leads of the small dot (see Supplementary Information for analysis of temperature dependence). Using the same scaling relation as above (equation (3)), the data again collapse onto a single (inverted) curve at low bias and temperature (Fig. 3h). Interestingly, we find the numerical prefactor of κ to be 0.25 in both 1CK regimes—precisely matching each other, but only roughly agreeing with our predicted value of 0.82.

Having established the existence of two distinct Kondo ground states—depending on the ratio J_{ir}/J_{fr}—we next demonstrate the tunability necessary to reach the symmetric 2CK state, $J_{ir} \approx J_{fr}$. By setting the tunnel coupling to the finite reservoir to an intermediate value, we can observe either zero-bias enhancement or zero-bias suppression (marked by white arrows in Fig. 4a), in both cases away from any charge degeneracy point of the finite reservoir (marked by a black arrow in Fig. 4a). This is expected[4], as the antiferromagnetic coupling to a reservoir depends not only on a tunnelling rate but also on the energy required to transfer an electron from the local site to that reservoir. Gate bp tunes that addition energy for the finite reservoir, modifying J_{fr} while keeping J_{ir} nearly constant. The region in (sp, bp)—we use sp to indicate the voltage on gate sp, and so on)—of suppressed conductance (red) grows rapidly with increasing coupling to the finite reservoir, as seen in Fig. 4b, c and d for c = −258, −256 and −254 mV, respectively.

The evolution of dI/dV_{ds} from zero-bias enhancement to zero-bias suppression as a function of the voltage on coupling gate c may be seen most clearly in Fig. 4e. We identify the curve for c = −260 mV as being very close to the 2CK symmetric point. At first sight, it is surprising that this curve does not display a clear cusp at low V_{ds} ($G \propto 1 - \text{const.}\sqrt{V_{ds}}$). In fact, conformal field theory predicts that at the 2CK symmetric point the differential conductance should depend quadratically on bias for $eV_{ds} < kT$, and should only cross over to $\sqrt{V_{ds}}$ behaviour at higher bias $eV_{ds} > 3kT$ (see green curve in Fig. 4f). Such a crossover is hard to see in a single plot of differential conductance versus bias. Instead, we combine the dependence of differential conductance on both bias and temperature in a scaling plot to produce compelling evidence for 2CK behaviour. The expected scaling form is somewhat different from that for 1CK[11,16,23]:

$$\frac{g(0, T) - g(V_{ds}, T)}{T^{\alpha_2}} = \kappa_2 Y\left(\frac{eV_{ds}}{kT}\right) \quad (4)$$

Here $\alpha_2 = 0.5$, $\kappa_2 = (g_0/2)(\pi/T_{2CK})^{\alpha_2}$, and

$$Y(x) = 1 - F_{2CK}(x/\pi) \approx \begin{cases} \frac{3}{\pi}\sqrt{x} - 1 & \text{for} \quad x \gg 1 \\ cx^2 & \text{for} \quad x \ll 1 \end{cases} \quad (5)$$

with $c \approx 0.0758$ (ref. 24). As with κ in the 1CK analysis, in practice we treat κ_2 as a free parameter for each set of gate voltages. F_{2CK}, the dependence of 2CK conductance on normalized bias, is found by conformal field theory[11,25,26].

Figure 4f shows that when we tune close to the 2CK symmetric point (c = −258 mV rather than −260 mV, due to a small shift in parameters), data at various temperatures and biases collapse onto each other and match the conformal field theory prediction (equation (4)), which is scaled vertically by κ_2. The horizontal axis is plotted as $(eV_{ds}/kT)^{0.5}$, to emphasize that $g(V_{ds}) \propto 1 - \text{const.}\sqrt{V_{ds}}$ for $eV_{ds}/kT \gg 1$.
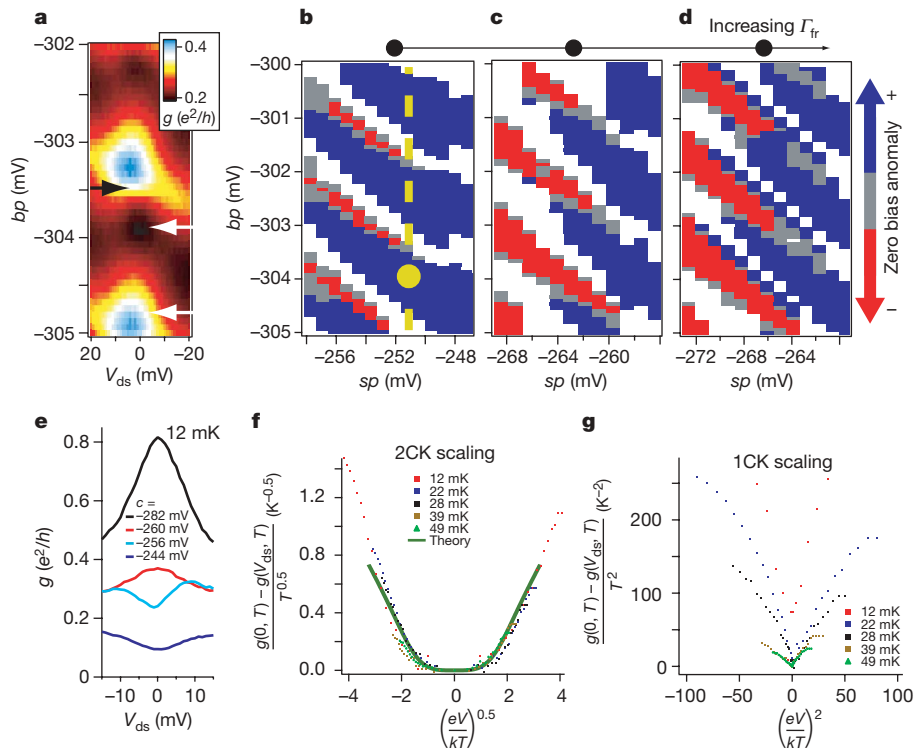
Figure 4 | Evidence for 2CK physics. a, Differential conductance as a function of bp and V_{ds} in the middle of a Kondo valley, with intermediate coupling to the finite reservoir, $c = -258$ mV. In contrast to Fig. 3c and g, here we observe both zero-bias enhanced and zero-bias suppressed conductance (marked by bottom and top white arrows, respectively) by fine-tuning bp. The black arrow marks the charge degeneracy point of the finite reservoir. **b–d,** At $c = -258$, -256 and -254 mV, conductance may be either enhanced or suppressed at zero bias, depending on the fine tuning of the electrostatic potentials of the small dot and finite reservoir with gates sp and bp, respectively. Red and blue indicate regions of suppressed conductance or enhanced conductance, respectively, while grey indicates relatively flat conductance around $V_{ds} = 0$. White regions are too close to charge degeneracy of dot or finite reservoir to diagnose Kondo-induced enhancement or suppression of conductance. Increased coupling to the finite reservoir expands the region of suppressed conductance (red). The

yellow dashed line in **b** shows the setting of sp used in **a**, and the yellow dot shows the approximate location in the charging hexagon of the symmetric 2CK point analysed in **f** and **g**. **e,** Differential conductance near zero bias evolves with coupling c from zero-bias enhancement to zero-bias suppression. For each curve, sp sets the small dot in the middle of a Kondo valley and bp sets the finite reservoir midway between two charge degeneracy points. **f, g,** Tuning bp near -304.7 mV (bottom white arrow in **a**), we observe that differential conductance depends on bias and temperature with $\alpha = 0.5$, consistent with 2CK (**f**) and inconsistent with 1CK (**g**, which attempts to apply scaling to exactly the same data, but with $\alpha = 2$.) In the Supplementary Information we show the converse, namely that in the 1CK region the 2CK scaling law does not fit, while the 1CK scaling does. A two-dimensional nonlinear fit to the data set used for **f** and **g** yields $\alpha_2 = 0.62 \pm 0.21$, consistent with $\alpha = 0.5$.

In Fig. 4g we show the same data scaled as would be appropriate for 1CK behaviour (equation (3)) instead of 2CK. As anticipated, this scaling fails completely: scaled data for different temperatures deviate from each other even near zero bias. A two-dimensional nonlinear fit to the data in Fig. 4f produces a value $\alpha_2 = 0.62 \pm 0.21$ (95% confidence limits), consistent with 2CK behaviour. Naively, we would expect 2CK behaviour to persist only up to $\{kT, eV_{ds}\} \approx (kT_K)^2/E_c \approx 1.7$ µeV (ref. 27). Empirically, 2CK persists to much higher energies: conductance follows the 2CK scaling form up to $V_{ds} = 15$ µeV, corresponding to $T = 180$ mK, even higher than T_K. Enhancement of 2CK energy scales has been predicted in our geometry in the presence of charge fluctuations[28], but is not expected to be so dramatic for our parameter values.

Here we have presented data demonstrating the existence of two independent 1CK states, along with a study of the associated 2CK state. Remarkably, the conductance of the symmetric 2CK state matches not only a simple power law but rather a complete theoretically calculated non-Fermi-liquid scaling function over a broad range of energy (equation (5), Fig. 4f). In future, it would be interesting to extend this scaling form theoretically and experimentally to cover the effects of a Zeeman field and slightly asymmetric coupling to the two reservoirs—for example, to quantitatively describe the family of curves in Fig. 4e. Finally, other parameter regimes of the same structure may show additional exotic behaviour[28,29].

1. Pruschke, T., Jarrel, M. & Freericks, J. Anomalous normal-state properties of high T_c superconductors: Intrinsic properties of strongly correlated electron systems? *Adv. Phys.* **44,** 187–210 (1995).
2. Cox, D. L. & Zawadowski, A. Exotic Kondo effects in metals: magnetic ions in a crystalline electric field and tunnelling centres. *Adv. Phys.* **47,** 599–942 (1998).
3. Hanson, R., Kouwenhoven, L. P., Petta, J. R., Tarucha, S. & Vandersypen, L. M. K. Spins in few-electron quantum dots. Preprint at ⟨http://arxiv.org/cond-mat/0610433⟩ (2006).
4. Oreg, Y. & Goldhaber-Gordon, D. Two-channel Kondo effect in a modified single electron transistor. *Phys. Rev. Lett.* **90,** 136602 (2003).
5. Wilson, K. G. The renormalization group: critical phenomena and the Kondo problem. *Rev. Mod. Phys.* **47,** 773–840 (1974).
6. Nozières, P. & Blandin, A. Kondo effect in real metals. *J. Phys.* **41,** 193–211 (1980).
7. Zawadowski, A. Kondo-like state in a simple model for metallic glasses. *Phys. Rev. Lett.* **45,** 211–214 (1980).
8. Goldhaber-Gordon, D. *et al.* Kondo effect in a single-electron transistor. *Nature* **391,** 156–159 (1998).
9. Grobis, M., Rau, I. G., Potok, R. M. & Goldhaber-Gordon, D. Kondo effect in mesoscopic quantum dots. Preprint at ⟨http://arxiv.org/cond-mat/0611480⟩ (2006).
10. Kondo, J. Resistance minimum in dilute magnetic alloys. *Prog. Theor. Phys.* **32,** 37–49 (1964).
11. Affleck, I. & Ludwig, A. W. W. Exact conformal-field-theory results on the multichannel Kondo effect: single-fermion Green's function, self-energy, and resistivity. *Phys. Rev. B* **48,** 7297–7321 (1993).
12. Cox, D. L. Quadrupolar Kondo effect in uranium heavy-electron materials. *Phys. Rev. Lett.* **59,** 1240–1243 (1987).

13. Seaman, C. L. *et al.* Evidence for non-Fermi-liquid behavior in the Kondo alloy $Y_{1-x}U_xPd_3$. *Phys. Rev. Lett.* **67,** 2882–2885 (1991).
14. Besnus, M. J. *et al.* Specific-heat and NMR of the Kondo system $YbPd_2Si_2$. *J. Magn. Magn. Mater.* **76–7,** 471–472 (1988).
15. Ralph, D. C. & Buhrman, R. A. Observations of Kondo scattering without magnetic-impurities — a point contact study of 2-level tunneling systems in metals. *Phys. Rev. Lett.* **69,** 2118–2121 (1992).
16. Ralph, D. C., Ludwig, A. W. W., von Delft, J. & Buhrman, R. A. 2-channel Kondo scaling in conductance signals from 2-level tunneling systems. *Phys. Rev. Lett.* **72,** 1064–1067 (1994).
17. Cichorek, T. *et al.* Two-channel Kondo effect in glasslike ThAsSe. *Phys. Rev. Lett.* **94,** 236603 (2005).
18. Glazman, L. I. & Raikh, M. E. Resonant Kondo transparency of a barrier with quasilocal impurity states. *JETP Lett.* **47,** 452–455 (1988).
19. Costi, T. A. & Hewson, A. C. Transport coefficients of the Anderson model via the numerical renormalization group. *J. Phys. Condens. Matter* **6,** 2519–2558 (1994).
20. Goldhaber-Gordon, D. *et al.* From the Kondo regime to the mixed-valence regime in a single-electron transistor. *Phys. Rev. Lett.* **81,** 5225–5228 (1998).
21. Göres, J. *et al.* Fano resonances in electronic transport through a single-electron transistor. *Phys. Rev. B* **62,** 2188–2194 (2000).
22. Majumdar, K., Schiller, A. & Hershfield, S. Nonequilibrium Kondo impurity: Perturbation about an exactly solvable point. *Phys. Rev. B* **57,** 2991–2999 (1998).
23. Hettler, M. H., Kroha, J. & Hershfield, S. Nonlinear conductance for the two channel Anderson model. *Phys. Rev. Lett.* **73,** 1967–1970 (1994).
24. Oreg, Y. & Goldhaber-Gordon, D. *Physics of Zero and One Dimensional Nanoscopic Systems* (Springer Series in Solid State Sciences, Springer, in the press).
25. Pustilnik, M., Borda, L., Glazman, L. & von Delft, J. Quantum phase transition in a two-channel-Kondo quantum dot device. *Phys. Rev. B* **69,** 115316 (2004).
26. von Delft, J., Ludwig, A. W. W. & Ambegaokar, V. The 2-channel Kondo model. II. CFT calculation of non-equilibrium conductance through a nanoconstriction containing 2-channel Kondo impurities. *Ann. Phys.* **273,** 175–241 (1999).
27. Florens, S. & Rosch, A. Climbing the entropy barrier: Driving the single- towards the multichannel Kondo effect by a weak Coulomb blockade of the leads. *Phys. Rev. Lett.* **92,** 216601 (2004).
28. Lebanon, E., Schiller, A. & Anders, F. B. Enhancement of the two-channel Kondo effect in single-electron boxes. *Phys. Rev. B* **68,** 155301 (2003).
29. Le Hur, K., Simon, P. & Borda, L. Maximized orbital and spin Kondo effects in a single-electron transistor. *Phys. Rev. B* **69,** 045326 (2004).
30. Haldane, F. D. M. Scaling theory of the asymmetric Anderson model. *Phys. Rev. Lett.* **40,** 416–419 (1978).

LETTERS

Chemical reduction of three-dimensional silica micro-assemblies into microporous silicon replicas

Zhihao Bao[1], Michael R. Weatherspoon[1], Samuel Shian[1], Ye Cai[1], Phillip D. Graham[1], Shawn M. Allan[1], Gul Ahmad[1], Matthew B. Dickerson[1], Benjamin C. Church[1], Zhitao Kang[1], Harry W. Abernathy III[1], Christopher J. Summers[1], Meilin Liu[1] & Kenneth H. Sandhage[1]

The carbothermal reduction of silica into silicon requires the use of temperatures well above the silicon melting point ($\geq$2,000 °C)[1]. Solid silicon has recently been generated directly from silica at much lower temperatures ($\leq$850 °C) via electrochemical reduction in molten salts[2,3]. However, the silicon products of such electrochemical reduction did not retain the microscale morphology of the starting silica reactants[2,3]. Here we demonstrate a low-temperature (650 °C) magnesiothermic reduction process for converting three-dimensional nanostructured silica micro-assemblies into microporous nanocrystalline silicon replicas. The intricate nanostructured silica microshells (frustules) of diatoms (unicellular algae) were converted into co-continuous, nanocrystalline mixtures of silicon and magnesia by reaction with magnesium gas. Selective magnesia dissolution then yielded an interconnected network of silicon nanocrystals that retained the starting three-dimensional frustule morphology. The silicon replicas possessed a high specific surface area ($>500\,\mathrm{m^2\,g^{-1}}$), and contained a significant population of micropores ($\leq$20 Å). The silicon replicas were photoluminescent, and exhibited rapid changes in impedance upon exposure to gaseous nitric oxide (suggesting a possible application in microscale gas sensing). This process enables the syntheses of microporous nanocrystalline silicon micro-assemblies with multifarious three-dimensional shapes inherited from biological[4–6] or synthetic silica templates[7–9] for sensor, electronic, optical or biomedical applications[10–13].

Silica diatom frustules were converted into MgO/Si-bearing replicas by the following reaction with gaseous Mg:

$$2Mg(g) + SiO_2(s) \rightarrow 2MgO(s) + Si(s) \qquad (1)$$

Images of a silica *Aulacoseira* diatom frustule, and of a reacted MgO/Si-bearing product, are shown in Fig. 1a and b, respectively. *Aulacoseira* frustules have a hollow cylindrical shape. Their end faces contain a circular hole with a protruding outer rim. These frustules were lined with rows of fine pores (10^2 nm in diameter) running along the cylinder length, and also contained narrow channels between intercalating fingerlike extensions. The MgO/Si composites (Fig. 1b) generated by reaction (1) retained the three-dimensional (3D) cylindrical morphology and nanoscale features (rows of fine pores, narrow channels) of the *Aulacoseira* frustules. X-ray diffraction (XRD) analysis (Fig. 2a) confirmed the presence of magnesia and silicon in the reacted specimens. The MgO/Si composite replicas were then immersed in a 1 M HCl solution for 4 h. XRD (Fig. 2b) and energy dispersive X-ray (Fig. 1e) analyses indicated that this treatment resulted in the selective and complete dissolution of magnesia.

The resulting silicon-based product (Fig. 1c) retained the 3D morphology and nanoscale features of the *Aulacoseira* frustules (see

Supplementary Fig. 1 for images of other diatom frustule replicas). The energy dispersive X-ray pattern in Fig. 1e contained a modest oxygen peak, which was associated with amorphous silica (silicon dioxide) formed during exposure of the silicon to the water-rich HCl solution. The replicas were then immersed in an ethanol-based hydrofluoric acid (HF) solution for 20 min. The absence of an oxygen peak in the energy dispersive X-ray pattern in Fig. 1f indicated that the silica had been completely and selectively dissolved by this HF
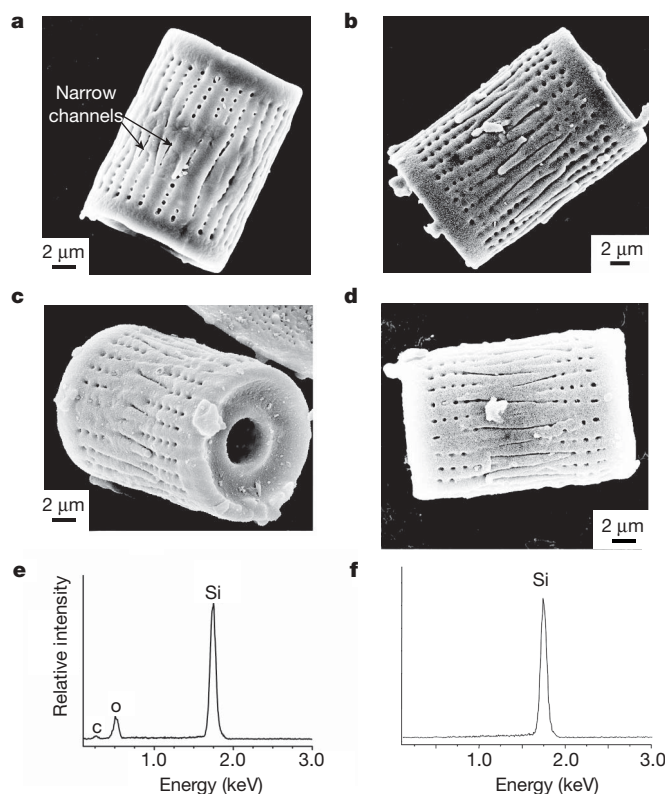

Figure 1 | Shape-preserving magnesiothermic reduction of silica diatom frustules. a, Secondary electron image of an *Aulacoseira* diatom frustule. **b**, Secondary electron image of a MgO/Si composite replica after reaction of an *Aulacoseira* frustule with Mg(g) at 650 °C for 2.5 h. **c**, Secondary electron image of a silicon-bearing replica produced by selective dissolution of magnesia from a MgO/Si replica in an HCl solution. **d**, Secondary electron image of a silicon replica after the HCl treatment and an additional treatment in a HF solution. **e, f**, Energy dispersive X-ray analyses obtained from silicon frustule replicas of the type shown in **c** and **d**, respectively.

[1]School of Materials Science & Engineering, Georgia Institute of Technology, Atlanta, Georgia 30332, USA.

treatment. X-ray photoelectron spectroscopy (Supplementary Fig. 2) and Fourier transform infrared spectroscopy (Supplementary Fig. 3) also confirmed the absence of silica in the specimens after the HF treatment. The replicas exposed to this HF treatment (Fig. 1d) retained the shape and fine features of the starting frustules.

A mixture of one mole of silicon with two moles of magnesia (the products of reaction (1)) corresponds to 34.9 vol.% silicon and 65.1 vol.% magnesia[14]. A uniform distribution of silicon and magnesia in such a product mixture should consist of co-continuous, interpenetrating silicon and magnesia phases[15]. The interconnectivity of both phases enabled the magnesia to be completely dissolved from the MgO/Si frustule replicas and the remaining silicon to be retained as a highly porous, but interconnected structure that preserved the starting frustule morphology. Nitrogen adsorption (Brunauer–Emmett–Teller, BET) measurements indicated that the specific surface areas of the starting silica frustules and of the MgO/Si composite replicas were only 1.65 m^2 g^{-1}, and 1.56 m^2 g^{-1}, respectively. After magnesia removal, however, the specific surface area increased to 541 m^2 g^{-1}. BJH (Barrett–Joyner–Halenda) analyses[16] of the nitrogen desorption curves indicated that, unlike the starting frustules or MgO/Si replicas, the silicon replicas possessed a significant population of micropores; that is, pores ≤2 nm in diameter comprised 30% of the cumulative volume occupied by all pores ≤125 nm in diameter (Supplementary Fig. 4).

The use of magnesium gas as a reducing agent enabled the reduction process to be conducted at a modest temperature, at which the formation of volatile silicon-bearing gas species (such as SiO(g), Si(g)) was suppressed[17]. As a result, the reduction reaction could be completed on and within the frustules (and not within the gas phase away from the frustule surfaces) to yield MgO/Si-bearing replicas. This modest 650 °C/2.5 h heat treatment, along with the formation of a continuous refractory magnesia phase intertwined with the silicon product, inhibited substantial coarsening and sintering of the silicon product, so that features as fine as a few tens of nanometres in the starting silica diatom frustules could be preserved in the silicon replicas after selective magnesia dissolution (see Supplementary Fig. 1). Although a few authors have previously examined the reaction of silica with magnesium gas, the shape preservation and microporosity of the liberated (magnesia-free) silicon product generated by such reaction were not examined[18–20].

Transmission electron images of cross-sections of a silicon frustule replica are shown in Fig. 3a–d. A corresponding selected area electron diffraction pattern is shown in Fig. 3e. Consistent with XRD analysis (Fig. 2b), electron diffraction analysis indicated the presence of silicon as the only crystalline phase throughout the replica cross-section. The fine sizes of the silicon crystals (Fig. 3b) detected within the wall of the converted frustule were also consistent with Scherrer analysis[21] of the XRD pattern in Fig. 2b, which yielded an average silicon crystallite size of 13 ± 2 nm. The high-resolution transmission electron image shown in Fig. 3c and the associated Fourier-filtered

version of this image in Fig. 3d both reveal lattice fringe images of adjacent silicon nanocrystals. The continuation of the lattice fringe patterns from these adjacent grains up to the grain boundary indicated that this grain boundary was free of an amorphous silica phase. Grain boundaries between other adjacent silicon nanocrystals were also found to be free of an amorphous phase. Some of the silicon crystals located at the outer surfaces of the silicon replicas were relatively large (~30 nm; Fig. 3b). A modified protocol was used to reduce further the average size of the crystals in the silicon replicas. The MgO/Si composite replicas were immersed in an aqueous solution of 1.1 M hydrogen peroxide and 2.2 M ammonium hydroxide at 70 °C for 0.5 h partially to oxidize the silicon grains (particularly the relatively coarse grains on the external replica surfaces). The specimens were then immersed in HF and HCl solutions to selectively dissolve the silica and magnesia, respectively. Scherrer analysis of the XRD pattern obtained from such silicon replicas yielded a reduced average silicon crystallite size of 8 ± 1 nm.

Planar (two-dimensional) microporous silicon has been used as a sensor for various gas species[11,22–24]. The high specific surface areas and open (accessible) structures of the present 3D silicon frustule replicas are attractive characteristics for use in sensitive and rapid gas detection. To evaluate such gas detection, a simple sensor based on a silicon frustule replica was fabricated (Fig. 4a). Platinum electrodes were used to connect the ends of this replica to gold pads (not shown) deposited on a silicon nitride substrate. A modest bias voltage (100 mV at 100 Hz) was applied across the electroded silicon frustule
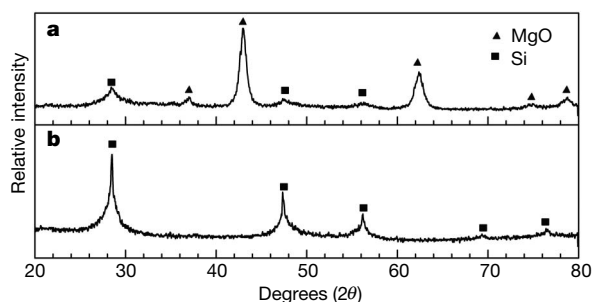

Figure 3 | Fine-scale structure of silicon replicas produced by magnesiothermic reduction of *Aulacoseira* diatom frustules. a, b, Low-magnification bright field and higher-magnification dark field transmission electron images, respectively, of an ion-milled cross-section of a silicon frustule replica. **c,** High-resolution transmission electron image of adjacent silicon nanocrystallites. **d,** The same image in **c** after Fourier filtering. **e,** Selected area electron pattern obtained from the silicon frustule replica shown in **a** and **b**.

Figure 2 | Phase content of diatom frustules after magnesiothermic reduction. a, XRD pattern revealing the presence of magnesia and silicon formed after exposure of *Aulacoseira* silica diatom frustules to magnesium gas at 650 °C for 2.5 h. **b,** XRD pattern obtained after selective dissolution of the magnesia from the MgO/Si composites in **a**.
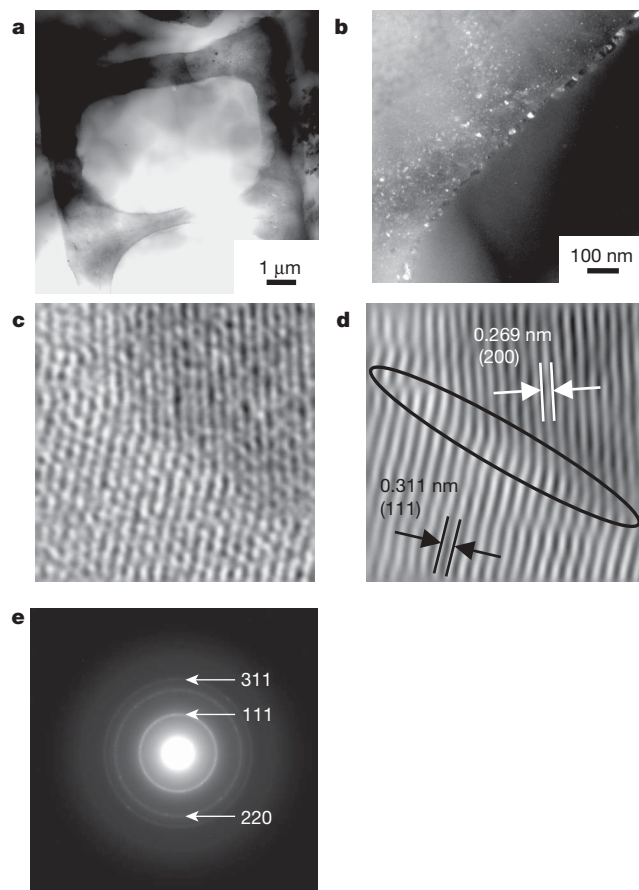
replica, and the change in impedance was evaluated upon exposure to flowing NO(g) in an Ar(g) carrier stream at 300 °C. The impedance of the silicon frustule sensor increased upon exposure to low (p.p.m.) concentrations of NO(g) (Fig. 4b). Furthermore, changes of only 1 p.p.m. in NO(g) concentration could be detected. The response (rise) and recovery (decay) times of this sensor were about 6 and 25 s, respectively, which are faster times than has been reported for planar porous silicon sensors exposed to comparable NO(g) concentrations[23]. Such a combination of rapid ($\leq$25 s) and sensitive (1 p.p.m.) detection of NO(g) from a microscale (minimally invasive) 3D silicon sensor operating with a bias voltage of only 100 mV (instead of several volts) has not previously been achieved, to our knowledge. Although further work remains to be conducted to optimize the packaging of such a sensor, the high sensitivity and rapid response rate of the silicon frustule replica demonstrate the benefits of such an open, porous, 3D structure with such high surface area.

Several authors have also reported the photoluminescence of microporous silicon in ultraviolet light, particularly after partial oxidation in water[25,26]. To evaluate such photoluminescence, the silicon frustule replicas were partially oxidized by immersion in water for 40 days. The photoluminescence of the 3D silicon frustule replicas was dramatically enhanced after such partial oxidation (Supplementary Fig. 5a and b). A fluorescence microscope image of such a photoluminescent silicon frustule replica is shown in Supplementary Fig. 5c.

The present work demonstrates that microporous assemblies of silicon nanocrystals with intricate and well-controlled 3D morphologies may be synthesized by the shape-preserving magnesiothermic reduction of microscale silica-based structures at only 650 °C. Biologically replicable silica-forming microorganisms, such as diatoms, can provide a wide variety of intricate 3D silica structures for this reduction process[4]. Sustained culturing of a given diatom species can yield significant quantities of 3D silica frustules of similar morphology. For example, the diatom biomass production rate of a single pilot-scale facility with 45 vertical bubble column photobioreactors can be in excess of 700 kg (dry weight) per annum[27], which corresponds to an annual silica frustule yield in excess of 70 kg (assuming a modest 10 wt% silica content in the dry cells[28]). The magnesiothermic reduction process is also amenable to scale-up (the current yield of the authors' laboratory-scale magnesiothermic reduction process is 100 mg h^{-1}, which is comparable to the 20–200 mg h^{-1} production rate of silicon nanoparticles via laser-induced pyrolysis of silane[29]). Hence, by applying the shape-preserving magnesiothermic reduction process to biologically or synthetically self-assembled silica templates, large quantities of similar 3D silicon nanocrystal assemblies with precisely controlled microscale shapes and nanoscale features (pores, channels and so on) may be synthesized for use in sensor, electronic, optical, biochemical, or other applications[10–13].

METHODS

Conversion of silica diatom frustules into microporous silicon replicas. Diatom frustules were spread evenly within a steel boat to form a 0.2-mm-deep powder bed. The frustule-bearing steel boat and magnesium granules were placed at opposite ends of a steel ampoule that was then welded shut. The Mg:SiO$_2$ molar ratio sealed within each ampoule was 2.5:1. Such ampoules were thrust into a tube furnace that had been preheated to 650 °C, and then held at this temperature for 2.5 h (at 650 °C, the oxygen and nitrogen partial pressures associated with the Mg/MgO and Mg/Mg$_3$N$_4$ equilibria are only 1.7×10^{-57} atm and 3.1×10^{-16} atm, respectively[17]). The reacted frustules were observed to contain three regions of different colour. The region located nearest the magnesium gas source was blue in colour. A black region and then a brown region were observed with increasing distance from the magnesium gas source. XRD analyses indicated that each region contained magnesia (magnesium oxide). The black and brown regions also contained silicon as a secondary product phase, whereas the blue region contained magnesium silicide, Mg$_2$Si (which is a blue-coloured intermetallic compound[14]). Residual, unreacted silica (in the form of cristobalite) was detected in the brown region. The MgO/Si composite material located in the black region of the reacted frustule bed was collected and immersed in a 1 M HCl solution (molar HCl:H$_2$O:EtOH ratio = 0.66:4.72:8.88) for 4 h at room temperature to selectively dissolve the magnesia. The specimens were then exposed to a HF solution (molar HF:H$_2$O:EtOH ratio = 1.05:1.11:6.45) to ensure that any silica formed during exposure to the aqueous HCl treatment was removed. The HCl treatment and the subsequent HF treatment were conducted within an argon atmosphere glove box. The argon in this glove box was obtained from an ultrahigh-purity tank (99.999% purity) and this gas was further scrubbed of oxygen with an oxygen gettering system. The oxygen partial pressure in this glove box was maintained at below 0.1 p.p.m. (as determined by an oxygen sensor).

Characterization of the microporous silicon replicas. The infrared spectra for FTIR measurements were recorded using a Bruker Equinox 55 spectrometer with a liquid N$_2$-cooled MCT detector. Measurements were made in diffuse reflectance mode using a Praying Mantis DRIFTS attachment from Harrick Scientific. Both the spectrometer and DRIFTS attachment were continuously purged with N$_2$ gas to minimize the background signal from atmospheric CO$_2$ and H$_2$O. The spectra obtained were averages of 128 scans recorded at 4 cm^{-1} resolution. The sample spectra, measured from samples consisting of ~10 vol.% silicon frustule replicas mixed with KBr powder, were referenced to a background of pure KBr powder. XPS analyses were conducted with a Kratos Axis-165 instrument (Kratos Analytical, Manchester, UK) using monochromatic Al Kα radiation. The specimens were transferred into the XPS instrument under an argon atmosphere. After removal from the HF-bearing solution, the specimens were allowed to dry in the argon atmosphere glove box. The frustule replicas were pressed onto 0.25-mm-thick indium foil and then sealed in plastic bags within the glove box. The sealed bags were then placed inside a glove bag that was, in turn, sealed around the evacuable antechamber of the XPS. After the glove bag was purged with ultrahigh-purity (99.999%) argon, the specimens were removed from the plastic bags and placed inside the antechamber. The antechamber was then evacuated for XPS analyses. Milling of the specimens within the XPS system was conducted with Ar$^+$ ions. XPS analyses were calibrated with respect to the carbon 1s excitation (285.0 eV).

For gas-sensing measurements, a single silicon frustule (*Aulacoseira*) replica was placed on a silicon nitride substrate and platinum electrodes were then applied to both ends of the replica with the aid of a focused ion beam instrument. A potentiostat was used to impose a small bias voltage (100 mV at 100 Hz) across the silicon frustule sensor, and to monitor the corresponding current passing through the sensor, during switching of the sample gas from pure argon to argon containing different concentrations of NO(g) at 300 °C.

The photoluminescence spectrum was obtained with a Spex 1000M spectrometer using an argon laser, with an emission wavelength of 275 nm (ultraviolet) and a power of 5 mW, as the excitation source. The laser was focused to a spot of ~1 mm in diameter on the sample. The photoluminescence measurements were
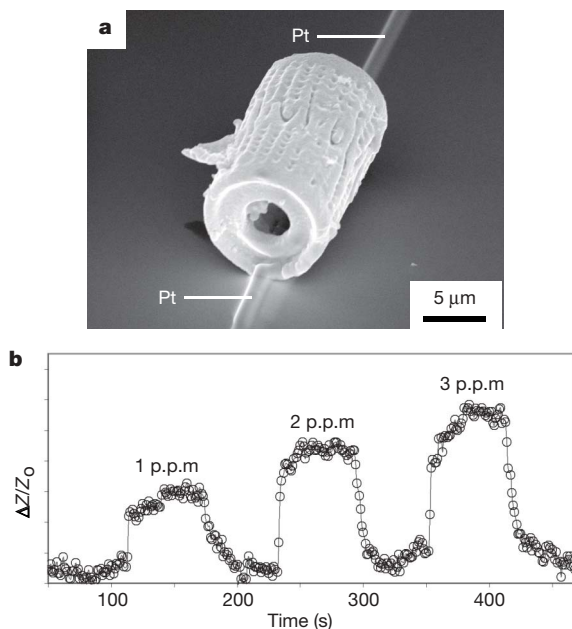

Figure 4 | Gas sensor based on a silicon frustule replica. a, Secondary electron image of an electroded microporous silicon frustule replica. **b,** Electrical response of this single silicon frustule sensor to NO(g). ΔZ is the impedance change upon exposure to NO(g), and Z_o is the sensor impedance in pure flowing argon. Measurements were obtained every second.

performed at room temperature and recorded with a GaAs photomultiplier tube with a cutoff wavelength of 920 nm. Fluorescence microscopy was conducted using a Zeiss Axiovert 200 microscope, with a fluorescence excitation of 450–490 nm.

Scanning electron microscopy was conducted with a Leo/Zeiss 1530 field emission scanning electron microscope operating at an accelerating voltage of 10 kV. Transmission electron microscopy was conducted with a JEOL 4000 EX microscope operating at an accelerating voltage of 400 kV.

1. Nagamori, M., Malinsky, I. & Claveau, A. Thermodynamics of the silicon-carbon-oxygen system for the production of silicon carbide and metallic silicon. *Metall. Trans. B* 17, 503–514 (1986).
2. Nohira, T., Yasuda, K. & Ito, Y. Pinpoint and bulk electrochemical reduction of insulating silicon dioxide to silicon. *Nature Mater.* 2, 397–401 (2003).
3. Yasuda, K., Nohira, T., Takahashi, K., Hagiwara, R. & Ogata, Y. H. Electrolytic reduction of a powder-molded SiO_2 pellet in molten $CaCl_2$ and acceleration of reduction by Si addition to the pellet. *J. Electrochem. Soc.* 152, D232–D237 (2005).
4. Round, F. E., Crawford, R. M. & Mann, D. G. *The Diatoms: Biology and Morphology of the Genera* (Cambridge Univ. Press, New York, 2000).
5. Aizenberg, J. *et al.* Skeleton of Euplectella sp.: structural hierarchy from the nanoscale to the macroscale. *Science* 309, 275–278 (2005).
6. Hall, S. R., Bolger, H. & Mann, S. Morphosynthesis of complex inorganic forms using pollen grain templates. *Chem. Commun.* 2784–2785 (2003).
7. Blaaderen, A., Ruel, R. & Wiltzius, P. Template-directed colloidal crystallization. *Nature* 385, 321–324 (1997).
8. Cha, J. N., Stucky, G. D., Morse, D. E. & Deming, T. J. Biomimetic synthesis of ordered silica structures mediated by block copolypeptides. *Nature* 403, 289–292 (2000).
9. Fan, H. *et al.* Rapid prototyping of patterned functional nanostructures. *Nature* 405, 56–60 (2000).
10. Cullis, A. G. & Canham, L. T. Visible light emission due to quantum size effects in highly porous crystalline silicon. *Nature* 353, 335–338 (1991).
11. Baratto, C. *et al.* Front-side micromachined porous silicon nitrogen dioxide gas sensor. *Thin Solid Films* 391, 261–264 (2001).
12. Shin, H.-C., Corno, J. A., Gole, J. L. & Liu, M. Porous silicon negative electrodes for rechargeable lithium batteries. *J. Power Sources* 139, 314–320 (2005).
13. Bengtsson, M. *et al.* Applications of microstructured porous silicon as a biocatalytic surface. *Phys. Status Solidi A* 182, 495–504 (2000).
14. *Powder Diffraction File* Card No. 27–1402 for Si, Card No. 45–946 for MgO, Card No. 35–0773 for Mg_2Si (International Center on Diffraction Data, Newtown Square, Pennsylvania); ⟨www.icdd.com⟩.
15. Bouvard, D. & Lange, F. F. Relation between percolation and particle coordination in binary powder mixtures. *Acta Metall. Mater.* 39, 3083–3090 (1991).
16. Barrett, E. P., Joyner, L. G. & Halenda, P. P. The determination of pore volume and area distributions in porous substances. I. Computations from nitrogen isotherms. *J. Am. Chem. Soc.* 73, 373–380 (1951).
17. Barin, I. *Thermochemical Data of Pure Substances* 3rd edn, pages 994, 1012, 1014, 1481, 1504, 1505 (VCH, Weinheim, 1995).
18. Wynnyckyj, J. R. & Rao, D. B. The mechanism of reduction of silica by magnesium vapor. *High Temp. Sci.* 8, 203–217 (1976).
19. Banerjee, H. D., Sen, S. & Acharya, H. N. Investigations on the production of silicon from rice husks by the magnesium method. *Mater. Sci. Eng.* 52, 173–179 (1982).
20. Sandhage, K. H. *et al.* Novel, bioclastic route to self-assembled, 3D, chemically tailored meso/nanostructures: shape-preserving reactive conversion of biosilica (diatom) microshells. *Adv. Mater.* 14, 429–433 (2002).
21. Cullity, B. D. *Elements of X-ray Diffraction* 101–102 (Addison-Wesley, Reading, Massachusetts, 1978).
22. Watanabe, K., Okada, T., Choe, I. & Sato, Y. Organic vapor sensitivity in a porous silicon device. *Sensors Actuators B* 33, 194–197 (1996).
23. Boarino, L. *et al.* NO_2 monitoring at room temperature by a porous silicon gas sensor. *Mater. Sci. Eng. B* 69–70, 210–214 (2000).
24. Seals, L., Gole, J. L., Tse, L. A. & Hesketh, P. J. Rapid, reversible, sensitive porous silicon gas sensor. *J. Appl. Phys.* 91, 2519–2523 (2002).
25. Balagurov, L. A., Leiferov, B. M., Petrova, E. A., Orlov, A. F. & Panasenko, E. M. Influence of water and alcohols on photoluminescence of porous silicon. *J. Appl. Phys.* 79, 7143–7147 (1996).
26. Gelloz, B., Kojima, A. & Koshida, N. Highly efficient and stable luminescence of nanocrystalline porous silicon treated by high-pressure water vapour annealing. *Appl. Phys. Lett.* 87, 031107 (2005).
27. Miron, A. S., Gomez, A. C., Camacho, F. G., Grima, E. M. & Chisti, Y. Comparative evaluation of compact photobioreactors for large-scale monoculture of microalgae. *J. Biotechnol.* 70, 249–270 (1999).
28. Schmid, A.-M. M., Borowitzka, M. A. & Volcani, B. E. in *Cytomorphogenesis in Plants* (ed. Kiermayer, O.) Cell Biology Monographs 8 63–97 (Springer, Vienna, 1981).
29. Li, X., He, Y., Talukdar, S. S. & Swihart, M. T. Process for preparing macroscopic quantities of brightly photoluminescent silicon nanoparticles with emission spanning the visible spectrum. *Langmuir* 19, 8490–8496 (2003).

Author Contributions Z.B., M.R.W., S.M.A., P.D.G., M.B.D. and K.H.S. conceived, developed and demonstrated the low-temperature magnesiothermic reduction and selective dissolution process. S.S. prepared and tested silicon replica gas sensors. G.A. cultured diatoms for conversion. Transmission electron microscope and BET analyses were conducted by Y.C. and B.C.C., respectively. H.W.A. and M.L. conducted and analysed FTIR measurements. Z.K. and C.J.S. conducted and analysed photoluminescence measurements. M.B.D. conducted fluorescence microscopy. Overall data analyses, project planning and paper preparation were largely conducted by Z.B. and K.H.S.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to K.H.S. (ken.sandhage@mse.gatech.edu).

LETTERS

Continental ice in Greenland during the Eocene and Oligocene

James S. Eldrett[1]†, Ian C. Harding[1], Paul A. Wilson[1], Emily Butler[1] & Andrew P. Roberts[1]

The Eocene and Oligocene epochs (~55 to 23 million years ago) comprise a critical phase in Earth history. An array of geological records[1–5] supported by climate modelling[6] indicates a profound shift in global climate during this interval, from a state that was largely free of polar ice caps to one in which ice sheets on Antarctica approached their modern size. However, the early glaciation history of the Northern Hemisphere is a subject of controversy[3,7–9]. Here we report stratigraphically extensive ice-rafted debris, including macroscopic dropstones, in late Eocene to early Oligocene sediments from the Norwegian–Greenland Sea that were deposited between about 38 and 30 million years ago. Our data indicate sediment rafting by glacial ice, rather than sea ice, and point to East Greenland as the likely source. Records of this type from one site alone cannot be used to determine the extent of ice involved. However, our data suggest the existence of (at least) isolated glaciers on Greenland about 20 million years earlier than previously documented[10], at a time when temperatures and atmospheric carbon dioxide concentrations were substantially higher.

The existence of Northern Hemisphere ice sheets can be demonstrated back to the middle Miocene (~15 million years ago), on the basis of proxy data and direct documentation of ice-rafted debris (IRD) from the Nordic seas[9–11]. There are few records of earlier Cenozoic glacial activity in the boreal realm because most marine sedimentary sections of this age suffer from extremely poor age control, slow sedimentation rates and widespread hiatuses. These problems are widely assumed to have resulted from increased ocean circulation strength and glacioeustatic sea-level fall associated with ice build-up on Antarctica[12]. Therefore, failure to document IRD from pre-mid-Miocene sediments in the Northern Hemisphere should not be interpreted to indicate that ice was restricted to Antarctica alone. To understand better Eocene-Oligocene palaeoceanographic and palaeoclimatic events in the high northern latitudes we have generated a new stratigraphy for three key Deep Sea Drilling Project (DSDP)/Ocean Drilling Program sites (ODP sites 338, 643 and 913; Fig. 1a) by calibrating dinocyst events to the geomagnetic polarity timescale[13]. The most complete Eocene to Oligocene succession is provided by ODP site 913 (75° 29.356' N, 6° 56.810' E, present
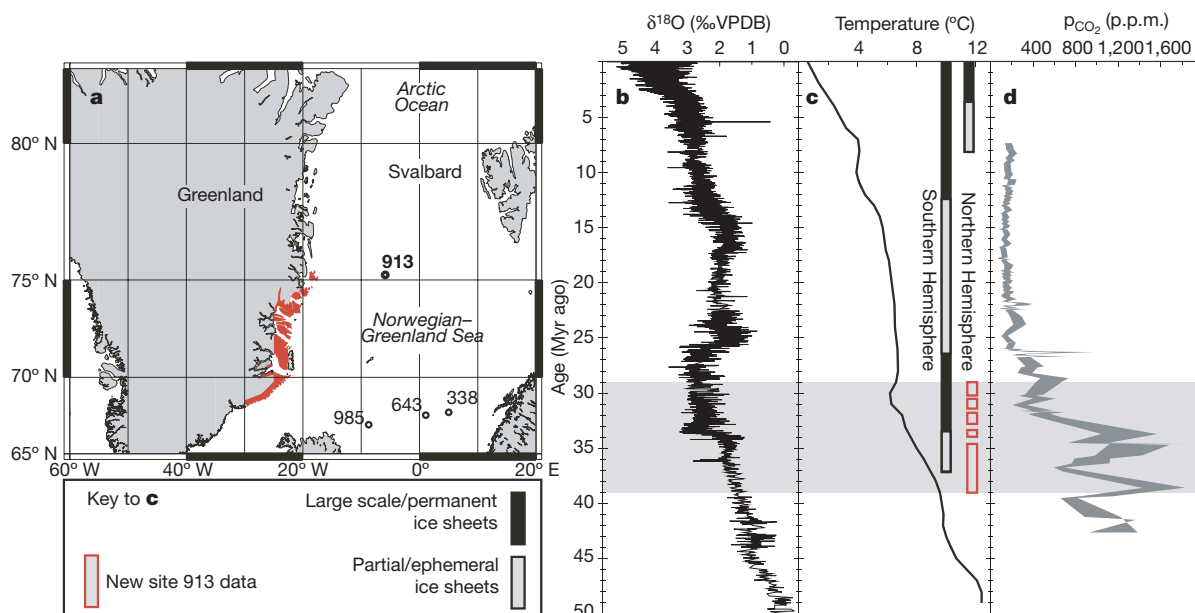

Figure 1 | Eocene–Oligocene palaeoclimate records and location of site 913. a, The location of site 913 (ODP Leg 151) and other sites referred to in the text. The area of East Greenland shaded in red indicates our inferred source area for the IRD: the Cretaceous–Cenozoic sedimentary basins and Palaeogene igneous outcrops from Kangerlussuaq in the south to Shannon Island in the north. **b–d**, The grey box indicates the stratigraphic interval studied here set against the benthic δ^{18}O curve from ref. 1 (**b**) and the Mg/Ca-based bottom water temperatures from ref. 2 (**c**), against which are set in black/grey the generally accepted chronology of southern and Northern Hemisphere Cenozoic glaciation as presented in ref. 1, and the view based on our data in red. VPDB, Vienna Pee-Dee belemnite standard. **d**, The CO_2 partial pressure curve from ref. 29.

[1]School of Ocean and Earth Science, National Oceanography Centre Southampton, University of Southampton, European Way, Southampton SO14 3ZH, UK. †Present address: Shell UK Ltd, 1 Altens Farm Road, Nigg, Aberdeen AB12 3FY, UK.

water depth ~3,300 m). All magnetic polarity chrons from chron C18n.2n to chron C11n are present with the exception of chron C13n, which we interpret to fall in the core gap that spans the Eocene–Oligocene boundary (Fig. 2a).

The Eocene–Oligocene section from ODP site 913 is carbonate-free, and is primarily composed of laminated biosiliceous ooze, overlain by clays and biosiliceous clays (Fig. 2b)[14]. Laminations are particularly pronounced from 471.6 m below sea floor (m.b.s.f.) to ~500 m.b.s.f. Initial shipboard investigations identified an isolated, macroscopic lithic clast at 453.16 m.b.s.f. but concluded that it was the result of down-hole contamination from the overlying Miocene glacial sequence, rather than an *in situ* glacially derived clast[14]. To test this shipboard interpretation, we undertook an investigation of the Eocene–Oligocene sequence at ODP site 913B. Our detailed core observations reveal the presence of *in situ* macroscopic clasts (up to 3.5 cm in length) from cores 27R to 21R (from the top of chron C18n.1n through to chron C17n.2-3n, and within chrons C15r, C12r and C11r: that is from ~38 to 30 Myr ago; Fig. 2a, b), which argue against the shipboard interpretation of drilling contamination. To determine the origin of these clasts, we undertook a microscopic examination of their surface features and made detailed observations of the cores in which they occur. We interpret the impact-induced deformation of laminae underlying clasts at 494.32 m.b.s.f. and 492.76 m.b.s.f. (C17n.3n, Fig. 3a, b) in this otherwise fine-grained pelagic sedimentary succession to be indicative of a dropstone origin. A 3.5-cm-long unfoliated gneissic clast at 492.9 m.b.s.f. (C17n.3n, Fig. 3c, d) reveals a prominent gouge with well-developed parallel striations. Other clasts display surficial microscopic conchoidal fractures (for example, a 3-cm-long vein quartz clast at 453.29 m.b.s.f.; C12r). These surface features are all consistent with glacial abrasion and crushing, and support an ice-rafted origin for these dropstones. Although glacially derived sediments deposited at the shelf edge by advancing continental glaciers could have been transported downslope by gravity flow, our dropstone-bearing interval lacks sedimentary structures indicative of gravity flow-induced deposition (for example, debrites, graded bedding, Bouma sequences, and so on). Organic-carbon-rich laminated sediments, characteristic of anoxia at

the sediment–water interface, are stratigraphically extensive in the studied material, and on the basis of all textural and sedimentary structure lines of evidence we rule out a mass flow transport mechanism for our dropstones.

To determine whether the dropstones are indicative of isolated or more pervasive ice-rafting, we analysed surface textures of microscopic grains, calculated mass accumulation rates (MARs) of different grain size fractions (>63, >125 and >250 μm), and determined the magnetic properties and whole-rock trace element geochemistry of the dropstone-bearing strata. Grain size samples were taken from 429.57 to 503.22 m.b.s.f., but sampling was not continued below core 28R owing to core disturbance and evidence of gravity flow deposits (for example, graded beds, sharp basal contacts, and so on). Scanning electron microscope analysis of >250 μm quartz grains from the interval yielding macroscopic dropstones indicates a high frequency of angular and high-relief grains, and surface morphological features (for example, striations, straight/arcuate steps and conchoidal fractures) that are diagnostic of mechanical breakage and sculpting of grain surfaces in glacial environments[5,11,15] (see Fig. 3e–l). These observations point towards pervasive ice rafting. Grain size fraction data can be used to help determine the mode of ice rafting. Analysis of surface sediment samples from the Arctic Ocean indicates that the >63 μm fraction is a minor component (≤1 wt%) of sediments rafted beyond the inner coastal zone by sea ice, but is a major constituent of sediments rafted by icebergs[16,17]. When downslope processes can be excluded, the presence of this grain size fraction has been taken as evidence for iceberg rafting[17–19]. On this basis, in the Eocene–Oligocene interval we studied, iceberg rafting appears to have played a much more important role than sea ice rafting (in ~90% of our samples the >63 μm size fraction formed >1 wt%; in ~75% of our samples it formed >5 wt%; and in ~10% of our samples it formed >30 wt%).

To provide comprehensive records of detrital mineral concentrations and to check whether there is a correspondence with IRD concentrations[20], we determined the mass magnetic susceptibility (χ) and the natural remanent magnetization (NRM) of the studied sediments. The magnetic properties of the sediments in the analysed
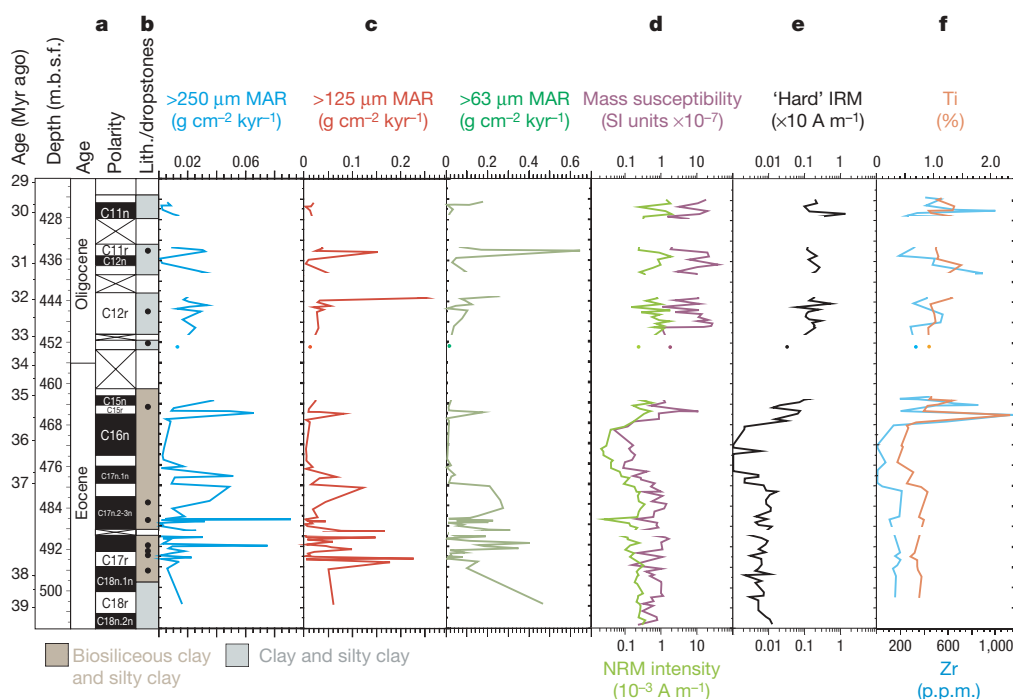

Figure 2 | Data from ODP site 913, Greenland Basin. a, Age model: the chronology for ODP site 913 is based on direct correlation to the geomagnetic polarity timescale reported in ref. 13. **b**, Lithology (Lith.) of the core, and positions of macroscopic dropstones (solid circles). **c**, Ice-rafted

debris mass accumulation rate (MAR) of the >250, >125 and >63 μm grain size fractions. **d**, **e**, Magnetic parameters are mass susceptibility (χ), NRM and 'hard' IRM, respectively (note the logarithmic scales). **f**, Trace-element abundances of zirconium and titanium (see Methods).

section vary considerably with depth (Fig. 2d, e). The magnetic mineral concentration is lowest in the interval with the highest biogenic component (C16n). These low χ and NRM values most probably result either from dilution associated with increased flux of biogenic material, or from diagenetic dissolution of detrital magnetic minerals attributable to increased organic carbon content in the biosiliceous interval[21]. Above C16n, there is a step-shift to much higher magnetic mineral contents (where χ and NRM reach maximum values). Fine-grained magnetite is the dominant magnetic mineral responsible for the magnetic signal at ODP site 913[13]. However, increased values of the 'hard' isothermal remanent magnetization (IRM) suggest a significant additional contribution from high-coercivity magnetic minerals (principally haematite) above C16n. The step from a magnetite-dominated signal to a magnetite–haematite signal may demonstrate different source terrains feeding ODP site 913 in the Eocene versus the Oligocene. Our X-ray fluorescence geochemical data series from ODP site 913 also records a prominent up-section shift from lower to higher values of Zr and Ti towards the top of C16n (Fig. 2f). This shift indicates an increased heavy and/or detrital mineral content that is consistent with a contemporaneous increase in terrigenous component[22] (see Supplementary Information, Fig. 2f).

Our data also allow us to help constrain the provenance of the IRD found at ODP site 913 and thereby to shed light on the location of early Northern Hemisphere ice formation. The $>63\,\mu m$ fraction in our samples is dominated by quartz grains, but basic igneous, sedimentary and metamorphic lithic clasts are also found, and these are most common in samples from the interval containing the greatest number of macroscopic dropstones (~C18n.1n to C17r, ~38 Myr

ago). The igneous clasts found throughout the Eocene sequence at ODP site 913, the haematite-dominated signal and the increase in titanium initiated within C16n (Fig. 2) are consistent with derivation from altered basalts. The provenance of this material may be the Palaeogene flood basalts of East Greenland, such as those on Hold with Hope[23] directly to the west of ODP site 913, but which have an extensive outcrop from Kangerlussuaq in the south to Shannon Island in the north (area marked in red in Fig. 1a). This hypothesis is supported by the presence of reworked early Eocene dinocysts above chron C16n at ODP site 913[13,24], which are compatible with those reported from the intra-basaltic Eocene sediments from East Greenland[25]. In addition, we have found dinocysts of late Cretaceous age[24], which are also known from sediments of Cenomanian to Maastrichtian age in East Greenland[26]. This observation is important because it allows us to rule out Svalbard as a potential source region, as post-Albian Cretaceous strata are absent from Svalbard[27].

Our data demonstrate that ice rafting to the Norwegian–Greenland Sea was active at least intermittently between 38 and 30 Myr ago at ODP site 913. Iceberg ice had an important role in delivering the IRD that we document, with East Greenland as the source (Fig. 1a). On the basis of our records alone, we cannot determine whether substantial ice-sheets existed on Greenland at this time, or whether the IRD originated from smaller, isolated glaciers. It remains to be seen, therefore, whether the events that we have documented would have significantly affected the isotopic composition of the contemporaneous ocean. Resolution of these questions has global implications because of the ongoing debate concerning Cenozoic ice budgets, the timing of the inception of Northern Hemisphere glaciation[2,3,7,9,12], and our
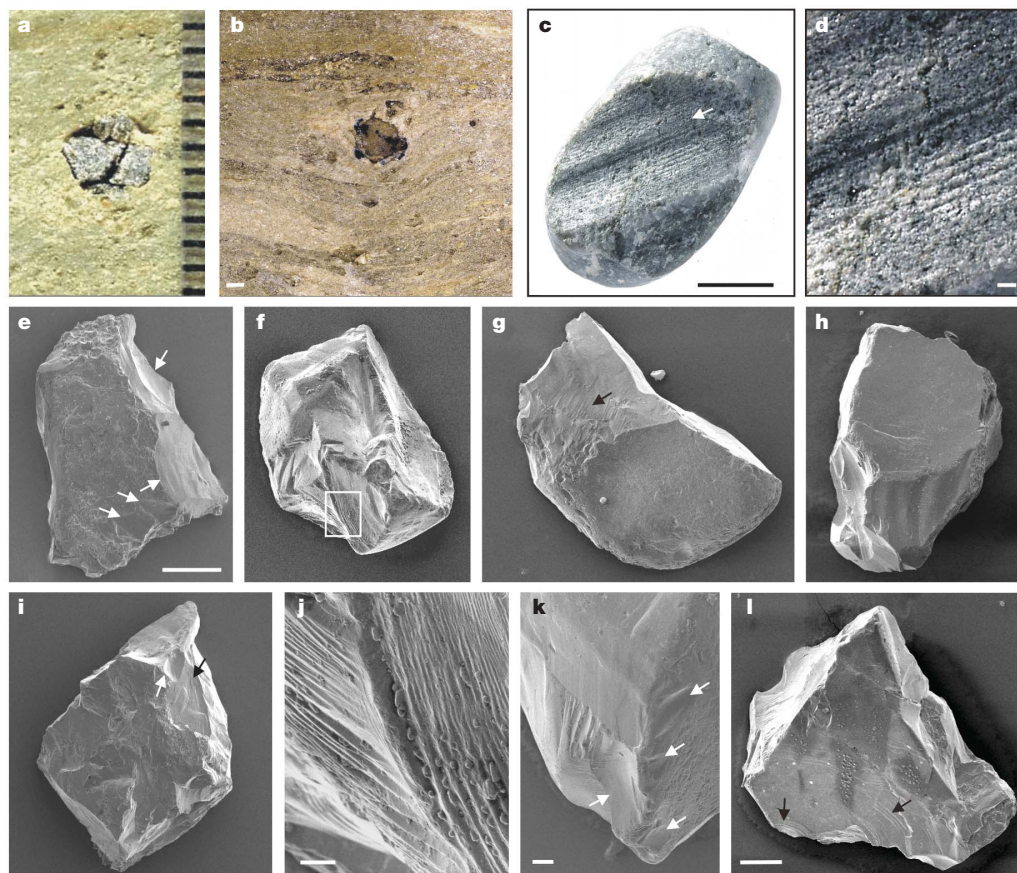

Figure 3 | Representative images of dropstones and quartz grains. a, b, *In situ* dropstones with impact-induced deformation of underlying laminae (**a**, 494.32 m.b.s.f.; **b**, 492.76 m.b.s.f.). **c, d,** Unfoliated gneissic dropstone from 492.9 m.b.s.f. with prominent gouge (white arrow in **c**) containing parallel striations. **d**, Detail of striations seen in **c**. **e–l**, Representative scanning electron microscope images of $>250\,\mu m$ quartz grains from 486.5 m.b.s.f., showing surface features indicative of glacial environments. White arrows in **e, i** and **k** indicate conchoidal fractures. Black arrows in **g, i** and **l** denote straight or arcuate stepped fractures, also shown in **f. j** is a magnified view of the area inside the white box in **f.** Scale bars: **a, b, d**, 1 mm (ruler scale in millimetres to the right of **a**); **c**, 1 cm; **e–i**, 100 μm (scale bar as in **e**); **j, k**, 10 μm; **l**, 100 μm.

understanding of Cenozoic atmospheric carbon dioxide levels (Fig. 1). Recent palaeoclimate model experiments generate substantial ice sheets in the Northern Hemisphere for the Eocene only in runs where carbon dioxide levels are lower (approaching the pre-anthropogenic level, $\sim 1 \times$ PAL) than suggested by proxy records (~ 7 to $2 \times$ PAL, Fig. 1d)[28,29]. Regardless, our data provide the first stratigraphically extensive evidence for the existence of continental ice in the Northern Hemisphere during the Palaeogene. This is about 20 Myr earlier than previously documented (mid-Miocene)[10], at a time when global deep water temperatures and, by extension, surface water temperatures at high latitude, were much warmer (Fig. 1c). By inference, our data strengthen the case for the existence of some ice in the Northern Hemisphere even earlier in the Cenozoic, as suggested by results from Integrated Ocean Drilling Program (IODP) site 302 (ref. 9), and for more extensive ice-sheets at the glacial maximum of the Eocene–Oligocene transition, as suggested by results from ODP site 1218 (ref. 3).

METHODS

IRD analysis. Grain size studies and surface texture analysis conducted on the coarse component of non-biogenic sediments from high latitudes are widely used to identify IRD[5,8,11,15–17]. Samples were freeze-dried and gently broken apart using a pestle and mortar. Hydrogen peroxide was added to remove organic matter. Once diluted, 10% hydrochloric acid was added to remove carbonate. Following the method of ref. 30, biogenic silica was removed by adding 2M potassium hydroxide to the samples, placing them in a water bath at 85 °C, followed by an ultrasonic water bath set to 100% and 80 °C and centrifuged for 10 min at 3,000 r.p.m. The remaining residue was sieved at 63, 125, 250 μm and weighed. The MAR of IRD was calculated using the following equation:

$$\text{IRD MAR} = (\%125 \,\mu m + \%250 \,\mu m) \times \text{sedimentation rate} \times \text{dry-bulk density}$$

Grain size composition was visually determined using a stereo-binocular microscope.

Photographic techniques. Core photographs and dropstone images were taken using standard digital photography at the Bremen ODP Core Repository, Germany, and at the National Oceanography Centre, Southampton (NOCS), UK. Digital scanning electron micrographs were made using secondary electron imagery with a LEO 1450 VP scanning electron microscope at NOCS. Individual grains were picked, mounted on glue pads on aluminium pin stubs and sputter-coated with gold to a thickness of ~ 20 nm. Secondary electron images were obtained using operating conditions of 15 kV, a nominal probe current of 120 pA and a working distance of 14 mm.

Rock magnetic techniques. Rock magnetic parameters were measured for ~ 100 samples at NOCS. Standard 7 cm^3 samples were given an anhysteretic remanent magnetization in a dc field comparable to the Earth's magnetic field (50 μT). The samples were then subjected to stepwise alternating field demagnetization at 10 mT increments, to a maximum peak field of 30 mT, and measured in a cryogenic magnetometer inside a magnetically shielded laboratory. All samples were given an IRM in a pulse magnetizer, with an inducing field of 1.2 T. The $\text{IRM}_{1.2\,T}$ was then demagnetized with a backfield of 300 mT. Magnetic susceptibility was measured using a Bartington Instruments MS2B dual-frequency magnetic susceptibility meter. All magnetic parameters used in this study are listed in the Supplementary Information. The S-ratio was calculated as follows:

$$S\text{-ratio} = -\text{IRM}_{-0.3\,T}/\text{IRM}_{1.2\,T}$$

$$\text{'Hard' IRM} = (\text{IRM}_{1.2\,T} - \text{IRM}_{-0.3\,T})/2$$

Geochemical techniques. Additional data were generated from discrete samples (5 to 10 g of dried sediment), determined using X-ray fluorescence spectroscopy on a Xepos sequential X-ray spectrometer system at Ichron Ltd, UK.

1. Zachos, J. C., Pagani, M., Sloan, L., Thomas, E. & Billups, K. Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science* **292**, 686–693 (2001).
2. Lear, C. H., Elderfield, H. & Wilson, P. A. Cenozoic deep-sea temperatures and global ice volumes from Mg/Ca in benthic foraminiferal calcite. *Science* **287**, 269–272 (2000).
3. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
4. Pekar, S. E., Hucks, A., Fuller, M. & Li, S. Glacioeustatic changes in early and middle Eocene (51–42 Ma): shallow-water stratigraphy from ODP Leg 189 Site 1171 (South Tasman Rise) and deep-sea δ18O records. *Bull. Geol. Soc. Am.* **117**, 1081–1093 (2005).
5. Strand, K., Passchier, S. & Näsi, J. Implications of quartz grain microtextures for onset of Eocene/Oligocene glaciation in Prydz Bay, ODP Site 1166, Antarctica. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **198**, 101–111 (2003).
6. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2. *Nature* **421**, 245–249 (2003).
7. Tripati, A., Backman, J., Elderfield, H. & Ferretti, P. Eocene bipolar glaciation associated with global carbon cycle changes. *Nature* **436**, 341–346 (2005).
8. Wolf-Welling, T. C. W., Cremer, M., O'Connell, S., Winkler, A. & Thiede, J. Cenozoic Arctic gateway paleoclimate variability: indications from changes in coarse-fraction composition (ODP Leg 151). In *Proc. ODP Sci. Res.* (eds Thiede, J., Myhre, A.M. & Firth, J.V.), **151**, 515–567 (ODP, College Station, Texas, 1996).
9. Moran, K. et al. The Cenozoic palaeoenvironment of the Arctic Ocean. *Nature* **441**, 601–605 (2006).
10. Winkler, A., Wolf-Welling, T. C. W., Stattegger, K. & Thiede, J. Clay mineral sedimentation in high northern latitude deep-sea basins since the Middle Miocene (ODP Leg 151, NAAG). *Int. J. Earth Sci.* **91**, 133–148 (2002).
11. Helland, P. E. & Holmes, M. A. Surface textural analysis of quartz sand grains from ODP Site 918 off the southeast coast of Greenland suggests glaciation of southern Greenland at 11 Ma. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **135**, 109–121 (1997).
12. Miller, K. G., Wright, J. D. & Fairbanks, R. G. Unlocking the icehouse: Oligocene-Miocene oxygen isotopes, eustasy, and margin erosion. *J. Geophys. Res.* **96**, 6829–6849 (1991).
13. Eldrett, J. S., Harding, I. C., Firth, J. V. & Roberts, A. P. Magnetostratigraphic calibration of Eocene-Oligocene dinoflagellate cyst biostratigraphy from the Norwegian-Greenland Sea. *Mar. Geol.* **204**, 91–127 (2004).
14. Myhre, A. M. et al. Site 913. *Proc. ODP Init. Rep.* **151**, 345–382 (1995).
15. Krinsley, D. H. & Doornkamp, J. C. *Atlas of Quartz Sand Surface Textures* (Cambridge Univ. Press, Cambridge, UK, 1973).
16. Clark, D. L. & Hanson, A. in *Glacial-Marine Sedimentation* (ed. Molnia, B. F.) 301–330 (Plenum, New York, 1983).
17. Nürnberg, D. et al. Sediments in Arctic ice: implications for entrainment, transport and release. *Mar. Geol.* **119**, 185–214 (1994).
18. Margolis, S. V. & Kennett, J. P. Cenozoic paleoglacial history of Antarctica recorded in subantarctic deep-sea cores. *Am. J. Sci.* **271**, 1–36 (1971).
19. Bond, G. et al. Evidence for massive discharges of icebergs into the North Atlantic ocean during the last glacial period. *Nature* **360**, 245–249 (1992).
20. Richter, T. O., Lassen, S., van Weering, T. C. E. & de Haas, H. Magnetic susceptibility patterns and provenance of ice-rafted material at Feni Drift, Rockall Trough: implications for the history of the British-Irish ice sheet. *Mar. Geol.* **173**, 37–54 (2001).
21. Karlin, R. & Levi, S. Diagenesis of magnetic minerals in recent hemipelagic sediments. *Nature* **303**, 327–330 (1983).
22. Hinrichs, J., Schnetger, B., Schale, H. & Brumsack, H.-J. A high resolution study of NE Atlantic sediments at station Bengal: geochemistry and early diagenesis of Heinrich layers. *Mar. Geol.* **177**, 72–92 (2001).
23. Upton, B. G. J., Emeleus, C. H. & Beckinsale, R. D. Petrology of the Northern East Greenland Tertiary flood basalts: evidence from Hold with Hope and Wollaston Forland. *J. Petrol.* **25**, 151–184 (1984).
24. Firth, J. V. Upper middle Eocene to Oligocene dinoflagellate biostratigraphy and assemblage variations in Hole 913B, Greenland Sea. *Proc. ODP Sci. Res.* **151**, 203–242 (1996).
25. Larsen, M., Heilmann-Clausen, C., Piasecki, S. & Stemmerik, L. in *Petroleum Geology: North-West Europe and global perspectives—Proceedings of the 6th Petroleum Geology Conference* (eds Doré, A. G. & Vinning, B. A.) 923–932 (The Geological Society, London, 2004).
26. Soper, N. J., Higgins, A. C., Downie, C., Matthews, D. W. & Brown, P. E. Late Cretaceous-early Tertiary stratigraphy of the Kangerdlugssuaq area, east Greenland, and the age of the opening of the north-east Atlantic. *J. Geol. Soc. Lond.* **132**, 85–104 (1976).
27. Harland, W. B. in *The Geology of Svalbard* (ed. Harland, W. B.) *Geol. Soc. Mem.* **17**, 363–387 (The Geological Society Publishing House, Bath, UK, 1997).
28. Deconto, R. M. & Pollard, D. Rethinking the Cenozoic record of ice volume: a modeling perspective on the relative contributions of Southern and Northern Hemispheres. *AGU Fall Meet. 2005* [CD] abstr. #PP52B–01 (American Geophysical Union, 2005); ⟨http://www.agu.org/⟩.
29. Pagani, M., Zachos, J. C., Freeman, K. H., Tipple, B. & Bohaty, S. Marked decline in atmospheric carbon dioxide concentrations during the Paleogene. *Science* **309**, 600–602 (2005).
30. Lyle, A. O. & Lyle, M. Determination of biogenic opal in pelagic marine sediments: a simple method revisited. *Proc. ODP Init. Rep.* **199**, 1–21 (2002).

LETTERS

Origin of avian genome size and structure in non-avian dinosaurs

Chris L. Organ[1], Andrew M. Shedlock[1], Andrew Meade[2], Mark Pagel[2] & Scott V. Edwards[1]

Avian genomes are small and streamlined compared with those of other amniotes by virtue of having fewer repetitive elements and less non-coding DNA[1,2]. This condition has been suggested to represent a key adaptation for flight in birds, by reducing the metabolic costs associated with having large genome and cell sizes[3,4]. However, the evolution of genome architecture in birds, or any other lineage, is difficult to study because genomic information is often absent for long-extinct relatives. Here we use a novel bayesian comparative method to show that bone-cell size correlates well with genome size in extant vertebrates, and hence use this relationship to estimate the genome sizes of 31 species of extinct dinosaur, including several species of extinct birds. Our results indicate that the small genomes typically associated with avian flight evolved in the saurischian dinosaur lineage between 230 and 250 million years ago, long before this lineage gave rise to the first birds. By comparison, ornithischian dinosaurs are inferred to have had much larger genomes, which were probably typical for ancestral Dinosauria. Using comparative genomic data, we estimate that genome-wide interspersed mobile elements, a class of repetitive DNA, comprised 5–12% of the total genome size in the saurischian dinosaur lineage, but was 7–19% of total genome size in ornithischian dinosaurs, suggesting that repetitive elements became less active in the saurischian lineage. These genomic characteristics should be added to the list of attributes previously considered avian but now thought to have arisen in non-avian dinosaurs, such as feathers[5], pulmonary innovations[6], and parental care and nesting[7].

Birds have the smallest genomes of all amniotes, with an average haploid genome size of only 1.45 pg of DNA or roughly 1.45 billion bases[8]. Birds are therefore a useful group in which to study the causes and consequences of small genome size and the mechanisms by which genomes contract. Small genome sizes may have been favoured by the demands of flight, explaining the constricted genome sizes seen within Aves[3,4]. Consistent with this suggestion is the finding that flightless birds have larger genomes than birds that fly[4], and that bats possess smaller genomes than do mammalian sister groups[9]. Some comparative analyses suggest that genome size reduction may have begun in lineages of non-avian reptiles sometime before the origin of flight in birds[10,11], implying that small genome size evolved first but that powered flight drove further genomic contractions.

Previous investigations of genome size evolution in amniotes have largely been constrained to living species, which constitute only about 1% of all species that have ever existed in this branch of life[12]. But including long-extinct animals into comparative genomics studies has the potential to reveal the origins and macroevolutionary trends of genomic novelties, as well as the timing of major genomic changes, more accurately than by focusing solely on living species[13–15]. Here we combine information from 31 dinosaur species

(some of which were extinct birds) with data on extant vertebrate species to characterize genome size and structure in extinct dinosaurs and evolutionary trends in amniote genome architecture.

In living organisms there is a well-known positive relationship between cell size and genome size[16]. Cell-size data are not generally available from extinct species, but it is possible to approximate osteocyte (bone-cell) size from fossilized bones. We sampled bone sections from 26 extant tetrapod species and fossilized bone from 31 extinct dinosaur species. We calculated osteocyte size directly from histological sections of bone by measuring the small pockets (lacunae) in the mineral matrix in which the bone cells resided during life (Fig. 1a). The distribution of osteocyte size in dinosaur bone is bimodal, with the two modes corresponding mainly to two major dinosaur groups, Ornithischia and Theropoda, with a single sauropod sample yielding an intermediate osteocyte size (Fig. 1b).

Using a phylogenetic tree and a bayesian implementation of the comparative method software Continuous[17,18], we found that osteocyte size predicted genome size in extant vertebrates well, explaining 59% of the variation (Fig. 1c) under a normal regression model and 32% of the variation under a generalized least-squares regression (GLS) model that corrects for the non-independence of data points arising from shared ancestry.

We then derived the posterior predictive distributions (see Methods and Supplementary Fig. 1) of genome size for the 31 extinct species of dinosaurs from this regression model. The posterior predictive distributions account for the phylogenetic relationships among taxa and for uncertainty about the true regression model. Except for *Oviraptor*, all of the inferred genome sizes for extinct theropods (average genome size of 1.78 pg, standard deviation s.d. = 0.3) fall within the narrow range of genome sizes for living birds (0.97–2.16 pg, haploid[8]), a result that follows from the restricted size range of theropod osteocytes.

Gaps among living taxa created by extinction might give the appearance of abrupt shifts in traits. This could be the case for genome size, which displays a punctuated distribution among extant vertebrates[10], even though genome evolution in vertebrates has been hypothesized to occur gradually and continuously[19]. Our results suggest that genome size evolution in dinosaurs was indeed abrupt, with a rapid reduction in size occurring between theropod and non-theropod dinosaurs (0.7 pg, or 28% difference in average genome size; phylogenetically corrected *t*-test, *P*-value = 0.008; Fig. 2), long before the origin of Aves and powered flight. If we consider *Herrerasaurus* a basal saurischian, as a recent study suggests[20], the decline in genome size is inferred to have begun abruptly earlier, at the base of Saurischia, denoted by a shift in genome size between the two primary clades of dinosaurs, ornithischians and saurischians (phylogenetically corrected *t*-test, *P*-value = 0.007). In either scenario, the theropods are characterized by a long period of relative

---

[1]Museum of Comparative Zoology, Harvard University, 26 Oxford Street, Cambridge, Massachusetts 02138, USA. [2]School of Biological Sciences, University of Reading, Reading, RG6 6AJ, UK.

Figure 1 | Primary histological data, distribution of average osteocyte-cell size in extinct dinosaur species, and regression lines derived from data on extant animals used to infer genome size and interspersed repetitive elements. Black lines in **c** and **d** indicate the average of a bayesian posterior distribution of GLS phylogenetically corrected regression lines and grey dotted lines indicate their 95% confidence intervals (see Supplementary Information). **a**, Histological micrograph of compact cortical secondary bone from a diaphyseal cross-section of an *Allosaurus fragilis* (extinct theropod dinosaur) radius. Circles indicate the type of cells that were measured and triangles indicate the type of cells that were not measured. **b**, Distribution of average osteocyte cell sizes in 31 extinct dinosaur species. Red and blue crosses indicate the median (vertical) and standard (horizontal) deviation for theropod and ornithischian species, respectively. **c**, The regression line (grey) is $\ln(\text{genome size}) = -1.8 + 0.56 \ln(\text{cell size})$, $r^2 = 0.59$ (P-value < 0.0001, H_0, $\beta 1 = 0$). The GLS line (black) is $\ln(\text{genome size}) = -0.87 + 0.36 \ln(\text{cell size})$, $r^2 = 0.32$ (Bayes factor = 9.9, H_0, $\beta 1 = 0$). **d**, The regression line (grey) is $\ln(\text{transposable elements}) = -3.7 + 1.69 \ln(\text{genome size})$, $r^2 = 0.58$ (P-value < 0.0001, H_0, $\beta 1 = 0$). The GLS line is $\ln(\text{transposable elements}) = -3.39 + 1.32 \ln(\text{genome size})$, $r^2 = 0.34$ (Bayes factor = 9.4, H_0, $\beta 1 = 0$).

genome-size stasis (phylogenetic *t*-test for difference between genome size of birds and non-avian theropods, *P*-value = 0.24; between extant birds and non-avian theropods, *P*-value = 0.38),

and the analysis suggests that the small, presumably gene-dense genomes in this clade have not changed substantially in size for the last 230 million years (Fig. 2).

Figure 2 | Haploid genome size (mean of posterior predictive distribution) mapped onto a phylogeny shows a reduction within saurischian dinosaurs, **the lineage to which birds belong.** Myr, million years.

By comparison, we observe no reduction in genome size in ornithischian dinosaurs compared with non-dinosaurian relatives. We estimate that extinct ornithischians (average genome size of 2.49 pg, s.d. = 0.42) had genome sizes within the range of extant non-avian reptiles, such as monitor lizards (*Varanus*, 2.05 pg) and *Gekko* (2.56 pg), but are more comparable to the genomes of closer archosaurian relatives, such as *Crocodylus* (3.21 pg) and *Alligator* (2.49 pg). We therefore hypothesize that the ornithischian genome was structured more similarly to living basal archosaurs than to birds.

The profile of the repetitive landscape in genomes of major amniote clades can provide insight into the molecular processes that have modulated the nearly 380-fold range of genome sizes observed in living vertebrates[21]. We estimated the percentage of the genome comprised of the dominant fraction of repetitive DNA (interspersed mobile elements) in 24 extant vertebrate species using RepeatMasker and over 119 megabases of bacterial artificial chromosome (BAC)-end and scaffold DNA sampled from online databases, supplemented with data from whole-genome sequencing projects. We then generated regression models against genome size as outlined above (Fig. 1d), in a way that simultaneously accounts for the uncertainty in using an inferred variable (genome size) to infer additional variables (repeat density) and uncertainty about the true regression model. We also attempted to reconstruct genome-wide GC content in dinosaurs, but found no significant correlations of GC content with genome size or cell size in extant species.

The estimated density of interspersed repetitive elements within the genomes of dinosaurs averages 12% (s.d. = 3.1) for ornithischian versus 8.4% (s.d. = 2.6) for non-avian theropod species. Comparisons of the inferred density of repetitive elements between theropod and non-theropod dinosaurs suggest an abrupt change in this trait along the theropod lineage (phylogenetically corrected *t*-test, $P = 0.02$). Again, the phylogenetic placement of *Herrerasaurus* does not change estimations of repeat density in dinosaurs, but when considered a basal saurischian the analysis suggests that repeat

density evolved abruptly earlier with the rise of Saurischia as opposed to Theropoda (the difference in retroelement density between saurischian and ornithischian dinosaurs in this scenario is significant: phylogenetically corrected *t*-test, *P*-value = 0.02). Yet there is no apparent difference in the inferred density of interspersed repetitive elements between birds and non-avian theropods (phylogenetic *t*-test for birds and non-avian theropods, *P*-value = 0.38; between extant birds and non-avian theropods, *P*-value = 0.1).

Genome size in eukaryotes is mediated primarily by proliferating and attenuating copies of retroelements over evolutionary time[21]. Analyses of mobile elements in extant sauropsid genomes suggest that chicken repeat 1 (CR1) long interspersed nuclear elements (LINEs) and related mammalian-wide interspersed repeat (MIR)-like short interspersed nuclear elements (SINEs) dominate the repetitive landscape and were probably active in the common archosaur ancestor some 250 million years ago[22]. Our results suggest that these ancient retroelements probably underwent different rates of lineage-specific expansion within ornithischians and saurischian dinosaurs, leading to a 50% difference in this genomic compartment between them, resulting in the repetitive landscape observed in living birds. Moreover, LINE elements in avian genomes appear to be almost exclusively represented by CR1 elements[23] though their activity appears to be nearly extinct in the chicken genome, on the basis of the paucity of full-length, intact CR1 elements within this host species[1]. Our estimates of the repetitive fraction of dinosaur genomes suggest that the reduction in CR1 activity began close to the base of carnivorous theropod or saurischian dinosaurs some 230 to 250 million years ago (Fig. 3). This model of reptilian genome evolution cannot be reconstructed from available bird–mammal whole-genome comparisons and emphasizes the value of incorporating long-extinct fossil evidence into phylogenomic studies.

Besides genes themselves, the architecture of whole genomes (genome size, gene number, chromosomal synteny, GC content, and the diversity of repetitive DNA elements) vitally links genome fluidity,

Figure 3 | Bar graph of interspersed repetitive elements in a range of extant vertebrate species and extinct dinosaur species inferred from genome size reconstructions and the correlation between repetitive element composition and genome size. Ornithischian dinosaur repetitive element densities are variable, especially within marginocephalians

(*Triceratops* and relatives), but as a whole occupy a range similar to crocodilians. Conversely, the inferred percentage of repetitive elements in non-avian theropod genomes is smaller than that of ornithischian species (phylogenetic *t*-test, *P*-value < 0.02) and falls within the range of living birds.

life history and reproduction[24]. In addition, because genome size in part affects cell size, it has direct consequences for the rate of cell division, transcriptional processes, and cellular respiration[25]. Consequently, it is thought that physiological demands may have constrained the evolution of genome size in endothermic vertebrates[10,26] by favouring smaller red blood cells that increase surface area to volume ratios, and therefore their ability to facilitate gas exchange (a constraint that mammals may have circumvented with enucleate red blood cells)[4,27]. Our results suggest that this component of endothermy in living birds may have originated early in the saurischian/theropod lineage with commensurate changes in genome size, a conjecture consistent with studies of dinosaurian growth physiology using bone palaeohistology[28,29]. The later secondary expansion of genome size in flightless birds[3] suggests that, even though flight and genome size may not have arisen together, the two may be functionally related, perhaps at a physiological level.

METHODS

See Supplementary Information for additional details.

Histology and genome size data. Histological slides from extant and extinct adult, sub-adult and juvenile animals were obtained from museum collections. Digital micrograph images were randomly assigned labels and cell size (volume) data were measured blind for a total of 1,423 lacunae (osteocytes). Only larger cells from compact bone tissue, both primary and secondary, were sampled to help ensure that cells were measured near their mid-axis, thereby minimizing variation caused by slicing cells on different planes during sectioning. Data on genome size were obtained from the Animal Genome Size Database[8].

Repetitive element data. Human, dog, rat, chicken and mouse repetitive genome data were taken from published genome assemblies. For species for which published data were not available, we used the software program RepeatMasker (http://www.repeatmasker.org) to summarize repeat density and GC content in BAC clone sequences covering over 119 Mb downloaded from GenBank and random regions along scaffolds from published genomes on the UCSC genome browser (http://genome.ucsc.edu/).

Trees and character matrices. Mesquite version 1.11 (http://mesquiteproject.org/) was used to create character matrices and phylogenetic topologies. Branch lengths were estimated from the fossil record supplemented with molecular studies. Various tree topologies (*Chrysemys* as a basal amniote and as the sister group to archosaurs; *Herrerasaurus* as a basal saurischian and as a basal theropod) and branch-length scaling were used to assess the robustness of our approach. As mentioned in the text, most results were robust to these changes in topology and branch lengths.

Regression model. The dependent variable y and an independent variable x measured on a sample of n species are assumed to be related by the regression model $y_i = \beta_0 + \beta_1 x_i + e_i$, where β_0 is the y-axis intercept and β_1 is the slope of the line relating the y_i to the x_i. The e_i are the random errors and the residual variance of the regression is given by the variance of the $e_i = \sigma^2$.

Represented as vectors, **y** and **x** contain the observed data across species and are studied in a GLS framework such that $p(\mathbf{y}|\mathbf{x}, m) \propto \exp\left\{ -[\mathbf{y} - (\beta_0 + \beta_1 \mathbf{x})]' \mathbf{V}^{-1} [\mathbf{y} - (\beta_0 + \beta_1 \mathbf{x})] \right\}$, where m denotes the regression model, and **V** is the expected variance–covariance matrix of the residual errors given by the phylogenetic tree describing the relationships among species[17,18].

We account for the uncertainty about the true regression model $p(\mathbf{y})$ by defining **y** as given above, and the integral is approximated by Markov chain Monte Carlo (MCMC) methods[30]. A Markov chain is constructed in which new values of the parameters of the regression model are proposed on successive iterations of the Markov chain. At each step in the chain, models are accepted or rejected by the Metropolis–Hastings algorithm[30]. The chain is allowed to run to convergence, after which it samples from the posterior distribution of $p(\mathbf{y}|\mathbf{x})$ and the posterior distribution of m. All MCMC chains ran for 5,010,000 iterations with a burn-in of 1,000.

Posterior predictive distributions. We obtain estimates for unknown values of the dependent variable $\tilde{\mathbf{y}}$ in a new sample for which **x** is known by using the posterior distribution of m from $p(\mathbf{y}|\mathbf{x})$. The probability of the unknown $\tilde{\mathbf{y}}$ can be written as $p(\tilde{\mathbf{y}}|\mathbf{x}) = \int_m p(\tilde{\mathbf{y}}|\mathbf{x}, m)p(m)\mathrm{d}m$, where the **x** values here correspond to species for which we wish to estimate the unknown $\tilde{\mathbf{y}}$. Our MCMC methods sample m and **e** together from their joint posterior distribution as derived in the initial regression modelling step, and assume that the e_i follow a multivariate normal distribution. These are combined to produce new values of $\tilde{\mathbf{y}} = \beta_0 + \beta_1 \mathbf{x} + \mathbf{e}$. The posterior predictive distribution of genome sizes considers

uncertainty in the parameters of the regression model as well as the inherent uncertainty of prediction summarized in σ^2.

Posterior predictive distributions of repetitive genetic elements are produced as above; let $\tilde{\mathbf{y}}$ denote the unknown values of these variables. The probability of these unknown values is $p(\tilde{\mathbf{y}}|\mathbf{x}) = \int_{\tilde{\mathbf{x}}} \int_m p(\tilde{\mathbf{y}}|\mathbf{x}, \tilde{\mathbf{x}}, m)p(m)p(\tilde{\mathbf{x}})\mathrm{d}m\mathrm{d}\tilde{\mathbf{x}}$. The Markov chain samples from posterior distributions of the model parameters (including residual error) derived from known data, and from the posterior distribution of the genome sizes of extinct dinosaurs. The posterior distribution of repetitive elements therefore considers uncertainty in the regression model and in the values of some of the predictors.

Model/method checking. We evaluated the adequacy of our regression models by generating simulated species data from a multivariate normal distribution of about $MN(\mu, \mathbf{V})$. We also individually removed several extant species (*Meleagris*, *Capra*, *Iguana* and *Alligator*); we then inferred their genome size from osteocyte data, and found them to be within 2% to 15% of the reported average values (see Supplementary Information for more details). These tests show that by adding back phylogenetic information, initially 'lost' during GLS phylogenetic correction, our inferences of genome size are actually more accurate than a standard regression model, despite reduced r^2 values for phylogenetically corrected correlations.

1. Hillier, L. W. *et al.* Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695–716 (2004).
2. Ellengren, H. The avian genome uncovered. *Trends Ecol. Evol.* **20**, 180–186 (2005).
3. Hughes, A. L. *Adaptive Evolution of Genes and Genomes* (Oxford Univ. Press, Oxford, 1999).
4. Hughes, A. L. & Hughes, M. K. Small genomes for better flyers. *Nature* **377**, 391 (1995).
5. Xu, X., Zhou, Z. & Prum, R. O. Branched integumental structures in *Sinornithosaurus* and the origin of feathers. *Nature* **410**, 200–204 (2001).
6. O'Connor, P. M. & Claessens, P. A. M. Basic avian pulmonary design and flow-through ventilation in non-avian theropod dinosaurs. *Nature* **436**, 253–256 (2005).
7. Horner, J. R. & Makela, R. Nest of juveniles provides evidence of family structure among dinosaurs. *Nature* **282**, 296–298 (1979).
8. Gregory, T. R. *Animal Genome Size Database* ⟨http://www.genomesize.com/⟩ (2005).
9. Van den Bussche, R. A. How bats achieve a small C-value: frequency of repetitive DNA in *Macrotus. Mamm. Genome* **6**, 521–525 (1995).
10. Waltari, E. & Edwards, S. V. Evolutionary dynamics of intron size, genome size, and physiological correlates in archosaurs. *Am. Nat.* **160**, 539–552 (2002).
11. Tiersch, T. R. & Wachtel, S. S. On the evolution of genome size of birds. *J. Hered.* **82**, 363–368 (1991).
12. Raup, D. M. *Extinction: Bad Genes or Bad Luck?* 3–21 (W. W. Norton & Company, New York, 1992).
13. Conway Morris, S. & Harper, E. Genome size in conodonts (Chordata): inferred variations during 270 million years. *Science* **241**, 1230–1232 (1988).
14. Masterson, J. Stomatal size in fossil plants: Evidence for polyploidy in majority of angiosperms. *Science* **264**, 421–423 (1994).
15. Thomson, K. S. & Muraszko, K. Estimation of cell size and DNA content in fossil fishes and amphibians. *J. Exp. Zool.* **205**, 315–320 (1978).
16. Gregory, T. R. The bigger the C-value, the larger the cell: Genome size and red blood cell size in vertebrates. *Blood Cells Mol. Dis.* **27**, 830–843 (2001).
17. Pagel, M. D. Inferring evolutionary processes from phylogenies. *Zool. Scr.* **26**, 331–348 (1997).
18. Pagel, M. D. Inferring the historical patterns of biological evolution. *Nature* **401**, 877–884 (1999).
19. Gregory, T. R. & Hebert, P. D. N. The modulation of DNA content: proximate causes and ultimate consequences. *Genome Res.* **9**, 317–324 (1999).
20. Langer, M. C. in *The Dinosauria* (eds Weishampel, D. B., Dodson, P. & Osmólska, H.) 25–46 (Univ. California Press, Berkeley, 2004).
21. Kazazian, H. H. Jr. Mobile elements: Drivers of genome evolution. *Science* **303**, 1626–1632 (2004).
22. Shedlock, A. M. *et al.* Phylogenomics of non-avian reptiles and the structure of the ancestral amniote genome. *Proc. Natl Acad. Sci. USA* **104**, 2767–2772 (2007).
23. Shedlock, A. M. Phylogenomic investigation of CR1 LINE diversity in reptiles. *Syst. Biol.* **55**, 902–911 (2006).
24. Petrov, D. A. Evolution of genome size: new approaches to an old problem. *Trends Genet.* **17**, 23–28 (2001).
25. Kozlowski, J., Konarzewski, M. & Gawelczyk, A. T. Cell size as a link between noncoding DNA and metabolic rate scaling. *Proc. Natl Acad. Sci. USA* **100**, 14080–14085 (2003).
26. Cavalier-Smith, T. in *The Evolution of Genome Size* (ed. Cavalier-Smith, T.) 104–184 (John Wiley & Sons, Chichester, 1985).
27. Szarski, H. Cell size and the concept of wasteful and frugal evolutionary strategies. *J. Theor. Biol.* **105**, 201–209 (1983).

28. Erickson, G. M., Curry-Rogers, K. & Yerby, S. A. Dinosaurian growth patterns and rapid avian growth rates. *Nature* **412**, 429–433 (2001).
29. Padian, K., Horner, J. R. & de Ricqlès, A. J. Growth in small dinosaurs and pterosaurs: the evolution of archosaurian growth strategies. *J. Vert. Paleontol.* **24**, 555–571 (2004).
30. Gilks, W. R., Richardson, S. & Spiegelhalter, D. J. in *Markov Chain Monte Carlo in Practice* (eds Gilks, W. R., Richardson, S. & Spiegelhalter, D. J.) 1–19 (Chapman and Hall, London, 1996).

LETTERS

A single type of progenitor cell maintains normal epidermis

Elizabeth Clayton[1], David P. Doupé[1], Allon M. Klein[2], Douglas J. Winton[3], Benjamin D. Simons[2] & Philip H. Jones[1]

According to the current model of adult epidermal homeostasis, skin tissue is maintained by two discrete populations of progenitor cells: self-renewing stem cells; and their progeny, known as transit amplifying cells, which differentiate after several rounds of cell division[1–3]. By making use of inducible genetic labelling, we have tracked the fate of a representative sample of progenitor cells in mouse tail epidermis at single-cell resolution *in vivo* at time intervals up to one year. Here we show that clone-size distributions are consistent with a new model of homeostasis involving only one type of progenitor cell. These cells are found to undergo both symmetric and asymmetric division at rates that ensure epidermal homeostasis. The results raise important questions about the potential role of stem cells on tissue maintenance *in vivo*.

The mammalian epidermis is organized into hair follicles interspersed with interfollicular epidermis (IFE), which consists of layers of keratinocytes (Fig. 1a)[4]. In IFE, proliferating epidermal progenitor cells (EPCs) are found in the basal cell layer. On commitment to terminal differentiation, basal cells exit the cell cycle and subsequently migrate into the suprabasal cell layers. Progenitors capable of generating both hair follicles and IFE lie in the hair-follicle bulge, but these cells appear to play no part in maintaining normal interfollicular epidermis[5–9]. Label-retaining studies show that IFE contains slowly cycling basal cells, which have been interpreted as stem cells that support clonal units of transit amplifying (TA) and differentiated cells[10,11], according to the stem/TA cell hypothesis. However, these studies are unable to reveal the dynamics of EPC behaviour during epidermal homeostasis. Previous genetic labelling studies to track the fate of proliferating cells have either required epidermal injury or have yielded too few labelled clones to permit quantitative analysis[12–15].

To track EPC fate in normal epidermis we have used inducible genetic marking to label a sample of cells and their progeny in adult mice. Animals transgenic for the tamoxifen-regulated mutant of *cre* recombinase (Ah*cre*ERT), expressed from the inducible CYP1A1 promoter, were crossed onto the R26EYFP/EYFP reporter strain, in which a conditional allele of enhanced yellow fluorescent protein (EYFP) is targeted to the Rosa26 locus (Supplementary Fig. S1a; refs 16, 17). In the resultant Ah*cre*ERT R26EYFP/wt heterozygotes, EYFP is expressed in a dose-dependent manner following transient expression of *cre* induced by a treatment with βNF and tamoxifen at 6–9 weeks of age (Supplementary Fig. S1). Cohorts of mice were culled for analysis at intervals after a single injection of the inducing drugs. Cells expressing EYFP and their labelled progeny were detected by confocal microscopy of wholemount epidermis[18]. At 2 days post-induction, only singly labelled cells were seen, at a frequency of 1 in 600 cells in the basal layer, indicating that the clusters of cells encountered at later time points are clones, each derived from a single progenitor cell (Fig. 1b and data not shown). Analysis of subsequent cohorts of mice

demonstrated clones that remained cohesive and expanded progressively in size (Fig. 1b, Supplementary Fig. S2). We scored clones that contained one or more basal cells; the observed clone-size distribution (that is, the total number of nucleated cells per clone) up to 6 weeks post-labelling, and the basal-layer clone-size distribution up to one year (see Methods) are shown in Fig. 2.

The density of labelled clones containing at least one basal-layer cell in tail epidermis rose from 2 days to a peak at 2 weeks after induction, as EYFP levels accumulated to detectable levels in all labelled cells. Clone numbers then fell to $7 \pm 2\%$ (mean $\pm$ s.d.) of the peak value by 3 months, and $3 \pm 2\%$ at one year; similar results were seen in back skin (data not shown). This decline was accompanied by the appearance of multi-cellular clones containing only suprabasal cells, consistent with clonal loss through differentiation (Supplementary Fig. S4). Analysis of the spatial distribution of IFE clones indicates that labelled clones are not replaced by unlabelled
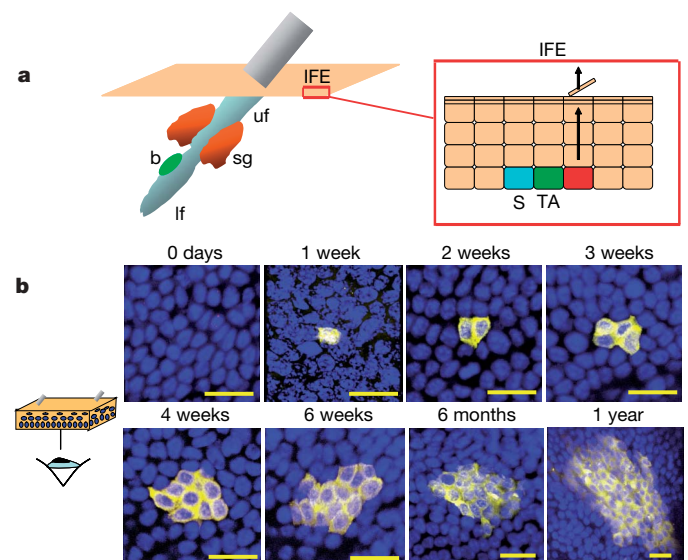
Figure 1 | *In vivo* clonal labelling of epidermal progenitor cells.
a, Organization of the epidermis. Hair follicles contain stem cells located in the bulge (b, green), with the potential to generate lower hair follicle (lf), sebaceous gland (sg, orange) upper follicle (uf) and interfollicular epidermis (IFE, beige). The schematic shows the organization of keratinocytes in the IFE, as proposed by the stem/TA cell hypothesis. The basal layer comprises stem cells (S, blue), transit amplifying cells (TA, dark green), and post-mitotic basal cells (red), which migrate out of the basal layer as they differentiate (arrows). **b**, Projected *Z*-stack confocal images of IFE wholemounts from Ah*cre*ERT R26EYFP/wt mice viewed from the basal surface at the times shown following induction. Yellow, EYFP; blue, DAPI nuclear stain. Scale bar, 20 μm.

[1]MRC Cancer Cell Unit, Hutchison-MRC Research Centre, Cambridge CB2 0XZ, UK. [2]Cavendish Laboratory, University of Cambridge, Madingley Road, Cambridge CB3 0HE, UK. [3]Cancer Research UK Cambridge Research Institute, Li Ka Shing Centre, Robinson Way, Cambridge CB2 0RE, UK.

clones migrating from hair follicles (Supplementary Fig. S5). Moreover, none of the labelled clones can derive from bulge stem cells because this region is not labelled (Supplementary Fig. S1 and Supplementary results).

Before attempting to interpret the clone fate data, it is necessary to assess the extent to which they are influenced by tissue growth or apoptosis. First, the rate of increase in epidermal surface area due to growth was low (estimated at less than 3.5% per month over the time course of the experiment), whereas apoptosis was undetectable in basal-layer cells (see Supplementary results and Supplementary Fig. S6). Furthermore, the number of basal-layer cells per unit area and the proportion of cycling cells (as assessed both by Ki67 and cdc6 immunostaining) showed no significant difference between 2-week and one-year samples. Both techniques of assessing the proportion of cycling cells gave similar results, as did flow cytometry: $22 \pm 3\%$ (mean $\pm$ s.d.) for Ki67; $24 \pm 4\%$ for cdc6; and $22 \pm 1\%$ for flow cytometry (see Supplementary Fig. S7)[19,20]. Finally, there was no significant difference between the proportion of cycling cells in the labelled and unlabelled cell populations, either at 5 days or one year post-induction (see Supplementary results and Supplementary Fig. S2). We therefore conclude that basal-layer cells labelled at induction are typical of the entire basal cell population, and that the observed clonal evolution is representative of the adult system in a state of homeostasis.

According to the stem/TA cell hypothesis, TA cells undergo a limited number of cell divisions followed by differentiation[21]. To test this prediction, we examined clones at 3 weeks, over 90% of which are lost by 12 weeks post-induction. Significantly, clones comprising three or more cells contained both basal and suprabasal cells, indicative of asynchronous terminal differentiation (Fig. 3a). Furthermore, the immunostaining of clones consisting of two basal cells reveals that a single cell division may generate either one cycling and one non-cycling daughter, or two cycling daughters, or two non-cycling daughters (Fig. 3b). This raises the question of whether there is asymmetric cell division within the basal plane as described in the *Drosophila* peripheral nervous system and zebrafish retinal precursors[22,23]. Three-dimensional imaging of wholemount epidermis revealed that only 3% of mitotic spindles lie perpendicular to the basal layer, indicating that, in contrast to embryonic epidermis, the vast majority of EPC divisions generate two basal-layer cells (Supplementary Fig. S8; refs 24, 25). The observation of asymmetric partitioning of numb protein (which marks asymmetric division in neural and myogenic precursors) in clones consisting of two basal cells suggests that planar-orientated asymmetric division also occurs in the epidermis (Fig. 3c)[26,27]. EPC behaviour thus differs substantially from that observed in committed precursors in other systems[28,29].

We next considered the behaviour of the long-lived clones that persist for over 3 months. Within the stem/TA cell hypothesis, the epidermis is organized into epidermal proliferative units comprising about ten basal cells supported by a single self-renewing stem cell[11]. If individual stem cells retain their self-renewal capacity, the stem/TA cell model predicts that the basal-layer clone-size distribution must become time-independent and characteristic of a single epidermal proliferative unit (see Supplementary theory[21]). Such behaviour is in stark contrast to the progressive increase in average clone size observed in the epidermis (Fig. 2).

Faced with this apparent contradiction, one could attempt to revise the stem cell/TA cell model, but staying within the general paradigm. This might include introducing the capacity for stem-cell ageing and/or migration[15]. Alternatively, one could try to exploit the range of experimental data to seek evidence for a new paradigm for epidermal homeostasis. Intriguingly, such evidence is found in the scaling properties of the observed clone-size distribution. Here we argue that the clone fate data are compatible with a model in which IFE is maintained by only one compartment of proliferating cells.
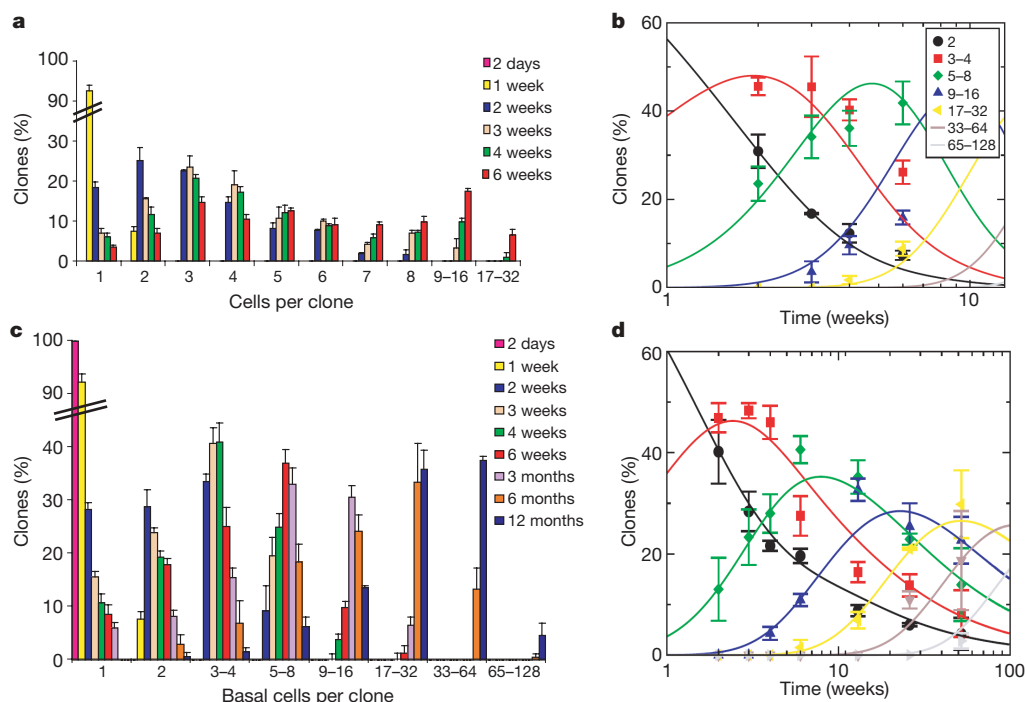
Figure 2 | Clone fate data. a, Distribution of clone size (total cells per clone) as a function of cell number, as measured at 2 days, 1, 2, 3, 4 and 6 weeks post-induction (error bars indicate s.e.m.). **c**, Distribution of basal cells per clone as a function of basal cell number, as measured at 2 days, 1, 2, 3 and 4 and 6 weeks, 3, 6 and 12 months post-induction (error bars indicate s.e.m.). **b**, **d**, Distribution of clone size (total cells per clone) (**b**) and basal cells per clone (**d**) as a function of time for different values of cell number (error bars indicate s.e.m.). Here we have aggregated clone sizes in ranges increasing in size in powers of two (see legend within figure). To eliminate possible ambiguities due to labelling efficiency, single cell clones are eliminated from the distribution in **b** and **d**, thereby removing the population of post-mitotic cells labelled at induction. We focus on time points of 2 weeks or more post-induction when EYFP levels have stabilized. Continuous curves show the behaviour of the proposed one-progenitor-cell model with a cell division rate of $\lambda = 1.1$ per week and a symmetric division ratio of $r = 0.08$ (see main text for details).
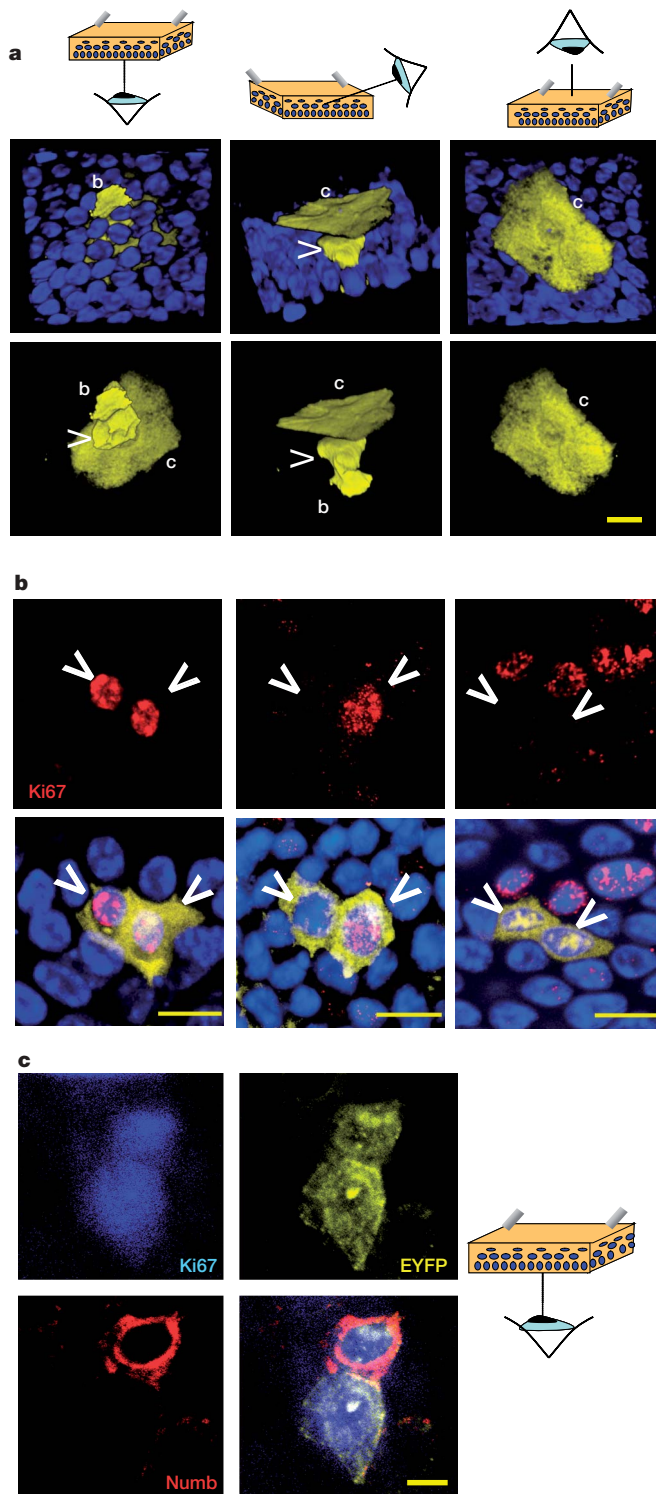
Figure 3 | Asymmetric cell fate in epidermal progenitors. a, Visualization of a three-cell clone exhibiting asynchronous terminal differentiation. Projected Z-stack images show one basal cell (labelled b), and two suprabasal cells: a cornified layer cell (labelled c), and a second suprabasal cell indicated by the arrowhead. Cartoon shows the angle of view. Upper panels: EYFP, yellow and DAPI, blue; lower panels are corresponding images with only EYFP shown. Scale bar, 20 μm. **b,** Visualization of two-cell clones (both cells basal, 3 weeks post-recombination), showing the different proliferative fates of the daughter cells of a single division, providing evidence for symmetric and asymmetric cell fate. Clones are viewed from the basal epidermal surface, stained for the proliferation marker Ki67 (red), DAPI (blue), and EYFP (yellow); arrowheads indicate position of EYFP-labelled cells. Three types of clone are shown, with two, one and zero Ki67 positive cells. Scale bar, 10 μm. **c,** Two-cell clone (both cells basal, 3 weeks post-recombination,

Whether this model should be considered as an extreme variant of the stem/TA cell hypothesis or a new concept is arguably a matter of semantics, a point we will return to later.

To identify the scaling behaviour, we define the clone-size distribution $P_n(t)$, describing the probability that a labelled progenitor cell develops into a clone with a total of n basal-layer cells at time t after induction. From this we can define the distribution of 'persisting' clones, that is, the distribution of labelled clones containing at least one basal-layer cell:

$$P_{n>0}^{\mathrm{pers}}(t) \equiv \frac{P_n(t)}{1 - P_0(t)}$$

With this definition, we show that (Fig. 4a and below), after an initial transient behaviour, the observed clone-size distributions are compatible with the simple scaling form:

$$P_{n>0}^{\mathrm{pers}}(t) = \frac{\tau}{t} f(n\tau/t) \qquad (1)$$

where τ denotes some constant timescale. From this striking observation, we deduce that, at long times, the average number of basal-layer cells within a persisting clone increases linearly with time, a behaviour inconsistent with the existence of long-lived cycling stem cells (see Supplementary theory). More significantly, the scaling indicates that long-time properties of clonal evolution are dictated by only one characteristic timescale τ, consistent with a simple model of clonal fate in which external factors, such as stem-cell ageing or skin injury, do not have a significant impact.

Taken together, all of our experimental observations and the scaling behaviour are consistent with a model of clonal fate involving only one type of EPC and just three adjustable parameters: the overall division rate λ of proliferating (labelled A-type) EPCs; the proportion of cell divisions that are asymmetric, $(1 - 2r)$; and the rate of transfer Γ of non proliferating (B-type) cells from the basal to the suprabasal layer (see schematic in Fig. 4b). To maintain a steady-state EPC population, the rates of symmetric cell division, $A \rightarrow A + A$ and $A \rightarrow B + B$, must be identical and equal to r. Finally, the observation that the basal-layer cell density remains constant leads to the additional constraint $\Gamma = \lambda \rho/(1 - \rho)$, where ρ denotes the proportion of proliferating cells in the basal layer, reducing the number of adjustable parameters to just two.

Defining $P_{n_A,n_B}(t)$ as the probability that a labelled clone involves n_A A-type and n_B B-type EPCs at time t after induction, its time-evolution is governed by the Master equation :

$$\frac{dP_{n_A,n_B}}{dt} = \lambda \Big\{ r \Big[(n_A - 1) P_{n_A-1,n_B} + (n_A + 1) P_{n_A+1,n_B-2} \Big] +$$

$$(1 - 2r) n_A P_{n_A,n_B-1} - n_A P_{n_A,n_B} \Big\} +$$

$$\Gamma \Big[(n_B + 1) P_{n_A,n_B+1} - n_B P_{n_A,n_B} \Big]$$

subject to the initial condition $P_{n_A,n_B}(0) = \rho\, \delta_{n_A,1}\delta_{n_B,0} + (1 - \rho)\, \delta_{n_A,0}\delta_{n_B,1}$. Although an exact analytical solution to this equation is unavailable, at times $t > 1/r\lambda$ the system enters an asymptotic regime where, defining $n = n_A + n_B$, we may show that the basal-layer clone-size distribution for persisting clones acquires the observed scaling form in equation (1) with $f(x) = e^{-x}$ and $\tau = \rho/r\lambda$; that is, the long-time properties of clonal evolution are dictated by the symmetric division rate, $r\lambda$. When combined with the experimentally inferred value for the fraction of proliferating basal-layer cells $\rho = 0.22$, a fit of the data to the asymptotic distribution (Fig. 4a, inset) identifies $r\lambda = 0.088 \pm 0.004$ per week.

viewed from the basal epidermal surface,) stained for the proliferation marker Ki67 (blue), numb (red) and EYFP (yellow), showing asymmetric distribution of numb, providing evidence for asymmetric cell fate resulting from a planar division. Scale bar, 5 μm.

a

b
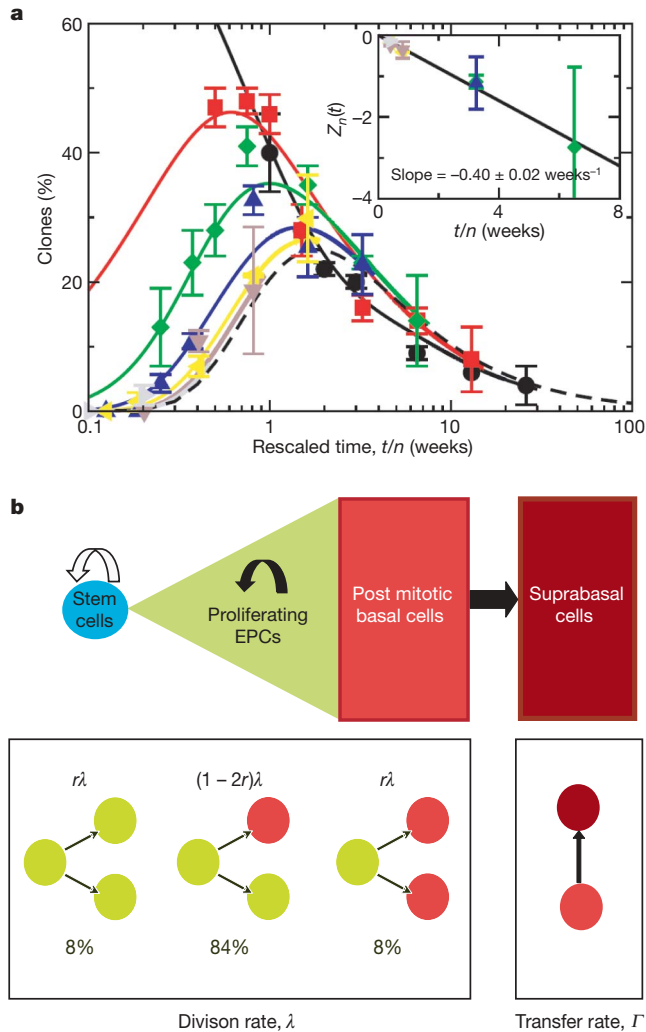
Figure 4 | Scaling and model of epidermal progenitor cell fate. a, The basal-layer clone-size distributions (see Fig. 2d legend for key) are replotted against the rescaled time coordinate, t/n (where n is taken as the upper limit for each distribution; for example, $n = 4$ for the range 3 to 4, and so on). We note that at long time points (>6 weeks), the data sets for different values of n converge onto a single curve (dashed line); that is, the probability of finding a labelled clone with a basal cell number in the range $n/2$ to n at time t post-induction is equal to that of finding a clone with a size in the range n to $2n$ at time $2t$. At shorter timescales, the transient behaviour dominates the distribution leading to a departure from simple scaling. This transient behaviour is very well described by the one progenitor cell compartment model (see Fig. 2 caption and main text). Making use of equation (1), we can identify the universal scaling curve for the grouped data as

$$G_n(t) \equiv \sum_{m > n/2}^{n} P_m^{\text{pers}}(t) = e^{-n\tau/2t} - e^{-n\tau/t}, \text{ where } \tau = \rho/r\lambda;$$ that is, $G_n(t/n)$ is

independent of n. Therefore, by plotting $Z_n(t) = 1/(2\ln[(1 - (1 - G_n(t))^{1/2})/2])$ against t/n (inset) at long times (>13 weeks) and large n (>4), the resulting slope may be used to infer $-1/\tau$. **b**, The single progenitor compartment model of epidermal homeostasis. A single population of EPCs (green triangle), with unlimited self renewal potential (filled arrow) maintain the epidermis. Post-mitotic cells in the basal layer (light red) transfer at a steady rate to the suprabasal compartment (dark red). The model proposes no role for stem cells (blue) in the steady state, but a quiescent population may play a role in growth or regeneration following injury (unfilled arrow). The left-hand box shows the proliferative characteristics of EPCs. Cycling cells are shown in green, post-mitotic basal cells in light red. The right-hand box shows the transfer of basal post-mitotic cells to the suprabasal layers (dark red). Taken together, these processes with their respective rates summarize the one-compartment model of homeostasis discussed in the text.

At times $t < 1/r\lambda \approx 11$ weeks, the transient behaviour of the basal-layer clone-size distribution dominates. In this regime, both the basal-layer and total clone-size distributions can be determined from a numerical integration of the corresponding Master equation. Taking $r\lambda$ from the asymptotic data dependence, a one-parameter fit of the basal-layer clone-size distribution to the experimental data obtains a good quantitative agreement over the entire one-year time course for $r = 0.08$ (Fig. 2d), that is, EPC division takes place at a rate of $\lambda = 1.1$ per week (a figure consistent with previous estimates[21]), with 84% of divisions resulting in asymmetric fate, while the cell transfer rate out of the basal layer takes place at a rate of $\Gamma = 0.31$ per week. With the same choice of parameters, the total clone-size distribution also shows a striking quantitative agreement with experiment (Fig. 2b). It should be noted that the slow accumulation of labelled clones over the 2 weeks following induction has no significant effect on the fit of the model (see Supplementary results and Supplementary Fig. S10.)

In conclusion, we have shown that the entire range of clonal fate data reported here is compatible with a model involving a single proliferating cell compartment in mouse tail skin epidermis. Technical limitations prevented us carrying out a similar quantitative analysis in back skin; although the changes in clone diameter observed at this site are in qualitative agreement with a single-compartment model, we cannot exclude alternative models with the available data. It is also important to note that, if present, a small quiescent population of stem cells would be undetectable in our analysis, but would be expected to be highly active in processes such as wound healing[8]. Previous models of epidermal homeostasis hypothesize the existence of a TA cell compartment, which undergoes a limited number of divisions. We show that tail epidermis is maintained by a single population of progenitor cells, which may undergo an unlimited number of divisions.

METHODS

Animals and sample preparation. All animal experiments were conducted as specified by Home Office Project Licence. The generation of Ah*cre*ERT and R26[EYFP/EYFP] mice has been described previously[16,17]. When Ah*cre*[ERT] R26[EYFP/wt] mice are treated with multiple doses of βNF and tamoxifen, a high level of recombination was seen in the upper hair follicle and IFE (Supplementary Fig. S1). The drug doses were titrated down to produce low-frequency labelling; a single intraperitoneal injection of 80 mg per kg β-naphthoflavone (Sigma-Aldrich) and 1 mg tamoxifen-free base (MP Biomedicals) dissolved in corn oil resulted in EYFP expression in approximately 1 in 600 basal cells of tail IFE and in 1 in 40 basal cells of back IFE, at 2 weeks post-induction[17]. No labelling was detected in the bulge region of the hair follicle and there was no background labelling in untreated Ah*cre*[ERT] R26[EYFP/wt] animals, even at 15 months of age (Supplementary Fig. S1c, d and data not shown). Epidermal wholemounts were prepared as described[18]. For analysis of back epidermis, 60-μm cryosections were used.

Immunostaining and imaging. Immunostaining of wholemounts was performed as described[18]. The following primary antibodies were used; anti-GFP conjugated to AlexaFluor488 or 555 (Molecular Probes), anti-Ki67 (Abcam), anti-numb (Abcam), anti-cleaved caspase 3 (Cell Signalling Technology) and anti-α-tubulin conjugated to FITC (Sigma). Secondary antibodies were from Molecular Probes. Confocal images are presented as Z-stack projections; 30–120 optical sections in 0.2–2 μm increments were rendered using Improvision Volocity software.

Analysis of clone size, number and proliferation. Data presented is a typical example of at least two experiments with at least three mice per time point.

Tail epidermis. The patterned organization of tail epidermis enabled definition of a unit area of tail IFE between adjacent rows of hair follicles, which measured $282,000 \pm 2,300$ μm[2] and contained $4,870 \pm 400$ (mean ± s.d.) basal-layer cells (Supplementary Fig. S5a). The mean number of labelled clones per unit area of tail IFE was assessed by counting all clones detected by confocal imaging of 50 unit areas in each of three mice at each time point, except at 6 and 12 months when 100 areas were counted. Clone size was determined by Z-stack confocal imaging of at least 50 clones containing at least one basal cell in each mouse at each time point. We note that after 6 weeks anucleate cornified layer cells appeared in labelled clones, making it impossible to count total cell numbers; hence, the total number of cells per clone was counted up to 6 weeks, while the total number of basal cells could be scored up to one year. The percentage of

Ki67-positive cells was determined by imaging at least 1,500 basal cells in multiple fields from at least three mice.

Back skin epidermis. 60-μm cryosections were analysed by optical sectioning. To determine clone number, all clones in three 10-mm lengths of epidermis were analysed in each mouse at time points up to 3 months; at 6- and 12-month time points at least 5 cm of epidermis was scored for each mouse. The maximum clone diameter, expressed as number of basal cells, was scored for at least 50 clones in each of three mice at each time point.

Apoptosis. Wholemounts of tail epidermis from 2 days, 1 week, 2 weeks, 3 weeks, 4 weeks and 6 weeks after induction were stained with an anti-cleaved caspase3/ AlexaFluor488 conjugate; positive caspase 3 staining was confined to catagen hair follicles. In further experiments, staining for cleaved caspase 3 and apoptotic bodies, visualized by DAPI, was examined in the basal layer in wholemounts from mice at 3 weeks and 6 months post-induction. As a positive control, epidermal wholemounts were irradiated with 160 mJ cm^{-2} ultraviolet (UV)C and analysed after a 16-hour incubation at 37 °C.

Analysis of mitotic spindle orientation. Wholemounts of DAPI-stained tail IFE were analysed by acquiring Z-stacks of all nuclei containing condensed chromosomes, as revealed by uniform, intense DAPI staining. Images were rendered in three dimensions, as above, and spindle orientation of all mitotic figures from metaphase to telophase was scored as described[22].

Analysis of _cre_ expression. _cre_^{ERt} messenger RNA levels after induction were analysed by quantitative polymerase chain reaction with reverse transcription (RT–PCR) of RNA prepared from tail epidermis using Trizol (Sigma). Primers used were 5′-CGTACTGACGGTGGGGAGAAT and 5′-CCCGGCAAAACAGG-TAGTTA, and the product was detected using SyBr Green. GAPDH mRNA was measured with a Taqman probe (Applied Biosystems).

Flow cytometry. A single-cell suspension was prepared from tail epidermis, using a 30 min incubation with Dispase II (Roche), followed by digestion with trypsin EDTA; this method separates the interfollicular epidermis from the dermis and lower hair follicles. After staining with biotin-conjugated anti-β1 integrin antibody (BD Biosciences/Pharmingen) and Alexa488-streptavidin (Molecular Probes), samples were fixed with paraformaldehyde, permeabilized with 0.1% saponin with 100 μg ml^{-1} RNase A and 50 μg ml^{-1} propidium iodide, and analysed on a BD Facscalibur flow cytometer, using propidium iodide channel pulse area/width gating to exclude cell doublets.

1. Lajtha, L. G. Stem cell concepts. _Differentiation_ **14**, 23–34 (1979).
2. Alonso, L. & Fuchs, E. Stem cells of the skin epithelium. _Proc. Natl Acad. Sci. USA_ **100** (suppl. 1), 11830–11835 (2003).
3. Braun, K. M. & Watt, F. M. Epidermal label-retaining cells: background and recent applications. _J. Invest. Dermatol. Symp. Proc._ **9**, 196–201 (2004).
4. Gambardella, L. & Barrandon, Y. The multifaceted adult epidermal stem cell. _Curr. Opin. Cell Biol._ **15**, 771–777 (2003).
5. Tumbar, T. _et al._ Defining the epithelial stem cell niche in skin. _Science_ **303**, 359–363 (2004).
6. Morris, R. J. _et al._ Capturing and profiling adult hair follicle stem cells. _Nature Biotechnol._ **22**, 411–417 (2004).
7. Levy, V., Lindon, C., Harfe, B. D. & Morgan, B. A. Distinct stem cell populations regenerate the follicle and interfollicular epidermis. _Dev. Cell_ **9**, 855–861 (2005).
8. Ito, M. _et al._ Stem cells in the hair follicle bulge contribute to wound repair but not to homeostasis of the epidermis. _Nature Med._ **11**, 1351–1354 (2005).
9. Claudinot, S., Nicolas, M., Oshima, H., Rochat, A. & Barrandon, Y. Long-term renewal of hair follicles from clonogenic multipotent stem cells. _Proc. Natl Acad. Sci. USA_ **102**, 14677–14682 (2005).
10. Mackenzie, I. C. Relationship between mitosis and the ordered structure of the stratum corneum in mouse epidermis. _Nature_ **226**, 653–655 (1970).
11. Potten, C. S. The epidermal proliferative unit: the possible role of the central basal cell. _Cell Tissue Kinet._ **7**, 77–88 (1974).
12. Ghazizadeh, S. & Taichman, L. B. Multiple classes of stem cells in cutaneous epithelium: a lineage analysis of adult mouse skin. _EMBO J._ **20**, 1215–1222 (2001).
13. Kameda, T. _et al._ Analysis of the cellular heterogeneity in the basal layer of mouse ear epidermis: an approach from partial decomposition _in vitro_ and retroviral cell marking _in vivo. Exp. Cell Res._ **283**, 167–183 (2003).
14. Ro, S. & Rannala, B. A stop-EGFP transgenic mouse to detect clonal cell lineages generated by mutation. _EMBO Rep._ **5**, 914–920 (2004).
15. Ro, S. & Rannala, B. Evidence from the stop-EGFP mouse supports a niche-sharing model of epidermal proliferative units. _Exp. Dermatol._ **14**, 838–843 (2005).
16. Srinivas, S. _et al._ Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. _BMC Dev. Biol._ **1**, 4 (2001).
17. Kemp, R. _et al._ Elimination of background recombination: somatic induction of Cre by combined transcriptional regulation and hormone binding affinity. _Nucleic Acids Res._ **32**, e92 (2004).
18. Braun, K. M. _et al._ Manipulation of stem cell proliferation and lineage commitment: visualisation of label-retaining cells in wholemounts of mouse epidermis. _Development_ **130**, 5241–5255 (2003).
19. Williams, G. H. _et al._ Improved cervical smear assessment using antibodies against proteins that regulate DNA replication. _Proc. Natl Acad. Sci. USA_ **95**, 14932–14937 (1998).
20. Birner, P. _et al._ Immunohistochemical detection of cell growth fraction in formalin-fixed and paraffin-embedded murine tissue. _Am. J. Pathol._ **158**, 1991–1996 (2001).
21. Potten, C. S. Cell replacement in epidermis (keratopoiesis) via discrete units of proliferation. _Int. Rev. Cytol._ **69**, 271–318 (1981).
22. Das, T., Payer, B., Cayouette, M. & Harris, W. A. _In vivo_ time-lapse imaging of cell divisions during neurogenesis in the developing zebrafish retina. _Neuron_ **37**, 597–609 (2003).
23. Gho, M. & Schweisguth, F. Frizzled signalling controls orientation of asymmetric sense organ precursor cell divisions in _Drosophila. Nature_ **393**, 178–181 (1998).
24. Lechler, T. & Fuchs, E. Asymmetric cell divisions promote stratification and differentiation of mammalian skin. _Nature_ **437**, 275–280 (2005).
25. Smart, I. H. Variation in the plane of cell cleavage during the process of stratification in the mouse epidermis. _Br. J. Dermatol._ **82**, 276–282 (1970).
26. Zhong, W., Feder, J. N., Jiang, M. M., Jan, L. Y. & Jan, Y. N. Asymmetric localization of a mammalian numb homolog during mouse cortical neurogenesis. _Neuron_ **17**, 43–53 (1996).
27. Conboy, I. M. & Rando, T. A. The regulation of Notch signaling controls satellite cell activation and cell fate determination in postnatal myogenesis. _Dev. Cell_ **3**, 397–409 (2002).
28. Smart, F. M. & Venkitaraman, A. R. Inhibition of interleukin 7 receptor signaling by antigen receptor assembly. _J. Exp. Med._ **191**, 737–742 (2000).
29. Temple, S. & Raff, M. C. Clonal analysis of oligodendrocyte development in culture: evidence for a developmental clock that counts cell divisions. _Cell_ **44**, 773–779 (1986).

Author Contributions Experimental work was performed by E.C., D.P.D. and P.H.J., project planning by P.H.J. and D.J.W., biophysical analysis by B.D.S. and A.M.K.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to P.H.J. (phj20@hutchison-mrc.cam.ac.uk).

nature

LETTERS

Hippocampal remapping and grid realignment in entorhinal cortex

Marianne Fyhn[1]*, Torkel Hafting[1]*, Alessandro Treves[1,2], May-Britt Moser[1] & Edvard I. Moser[1]

A fundamental property of many associative memory networks is the ability to decorrelate overlapping input patterns before information is stored[1–5]. In the hippocampus, this neuronal pattern separation is expressed as the tendency of ensembles of place cells[6] to undergo extensive 'remapping' in response to changes in the sensory or motivational inputs to the hippocampus[7–13]. Remapping is expressed under some conditions as a change of firing rates in the presence of a stable place code ('rate remapping')[14], and under other conditions as a complete reorganization of the hippocampal place code in which both place and rate of firing take statistically independent values ('global remapping')[14]. Here we show that the nature of hippocampal remapping can be predicted by ensemble dynamics in place-selective grid cells in the medial entorhinal cortex[15,16], one synapse upstream of the hippocampus. Whereas rate remapping is associated with stable grid fields, global remapping is always accompanied by a coordinate shift in the firing vertices of the grid cells. Grid fields of co-localized medial entorhinal cortex cells move and rotate in concert during this realignment. In contrast to the multiple environment-specific representations coded by place cells in the hippocampus, local ensembles of grid cells thus maintain a constant spatial phase structure, allowing position to be represented and updated by the same translation mechanism in all environments encountered by the animal.

Hippocampal place representations may be derived from a metric representation of space in the medial entorhinal cortex (MEC)[15–20]. A key cell type of the MEC is the grid cell, the spatial receptive fields form a regularly tessellating triangular pattern that spans the complete environment covered by the animal[16]. The periodic firing fields of neighbouring grid cells are offset relative to each other in an apparently random manner, such that all locations within the repeating unit of the triangular lattice are represented by any sufficiently large local cell ensemble[16]. To determine how neural activity in grid cells is organized at the ensemble level and how their organization might influence the formation of representations downstream in the hippocampus, we recorded the simultaneous activity of multiple grid cells in layer II of MEC in rats (*Rattus norvegicus*) under conditions in which place cells in the hippocampus undergo global remapping or rate remapping[14] (Fig. 1, Supplementary Figs 1 and 2, Supplementary Methods and Supplementary Tables 1–3). Cells were recorded simultaneously from MEC and area CA3 in 5 out of 15 rats with MEC electrodes. Global remapping was induced by two different procedures, one in which the rat was tested in the same location in either a square or a circular enclosure[7,9,21,22] (Fig. 1a) and another in which the animal alternated between similar square boxes in two different rooms[6,13,14] (Fig. 1b). These protocols induced pronounced changes in firing rate and firing location throughout the place cell population in all experiments in which CA3 cells were recorded (7 rats; 15

experiments; Fig. 1d–g). Rate remapping was induced by changing the walls of a square recording enclosure from a three black and one white configuration to a one black and three white configuration[14] (Fig. 1c). This induced consistent rate remapping in all seven experiments in which CA3 cells were recorded (five rats; Fig. 1d, e, h).

With distinct protocols for inducing global remapping and rate remapping in the hippocampus, we examined the collective neuronal firing patterns of grid cells in MEC in 21 experiments with 5 or more grid cells recorded simultaneously from the dorsocaudal part of MEC. Population dynamics in entorhinal grid cells was found to be strongly predictive of the type of remapping induced in the hippocampus. When global remapping was evoked by replacing the square box with the circular box, the cross-correlograms for individual grid cells (Fig. 2b, top) as well as simultaneously recorded ensembles (Fig. 2b, bottom) formed triangular grid-like patterns, suggesting that grid spacing, grid orientation and spatial phase distribution were preserved between the conditions (Fig. 2c, d). However, in all seven experiments using this protocol, the peak of the entorhinal cross-correlogram was offset from the origin when trials in different environments were compared. The offset was similar in all simultaneously recorded grid cells (Supplementary Figs 3–5). Between experiments, the displacement ranged from 10.0 to 23.8 cm, with the upper parts of the range corresponding to nearly one-half of the grid spacing at the recording location (red vectors in Fig. 2e; Supplementary Table 2 and Supplementary Figs 3c and 6). The displacement was significantly larger than between repeated trials in the same box (range 0.1–2.3 cm; blue vectors in Fig. 2e; $Z = 2.4$, $P < 0.01$, one-tailed Wilcoxon signed-ranks test). The direction of the offset was uniformly distributed between experiments (Fig. 2e; $r = 0.34$, $P > 0.05$, Rayleigh test for randomness). When cells were recorded from the same location over several days, the distance and the direction of the grid shift remained constant (Supplementary Fig. 3d). In three out of seven animals, the grid also showed a weak expansion from the square to the circle. The expansion was coherent among simultaneously recorded cells (Supplementary Fig. 5a and Supplementary Table 2). Cross-correlograms for simultaneously recorded CA3 cells were essentially flat (Fig. 2f and Supplementary Fig. 6), as would be expected if both firing location and firing rate are orthogonalized in CA3. These observations indicate that during global remapping in the hippocampus, the map of grid cells in the MEC realigns with changes in the environment without losing its intrinsic spatial phase structure.

To determine whether ensembles of grid cells remain coherent with more substantial changes in the external input, we compared the spatial phase relationship of simultaneously recorded grid cells during foraging in rooms with different distal landmarks (rooms A and B). Although the grid structure of individual cells was maintained across rooms, the periodicity was not immediately visible in

[1]Centre for the Biology of Memory, Norwegian University of Science and Technology, NO-7489 Trondheim, Norway. [2]Cognitive Neuroscience Sector, SISSA International School for Advanced Studies, I-34014 Trieste, Italy.
*These authors contributed equally to this work.

the population cross-correlograms (Fig. 3a–c). However, peaks became apparent when one stack was rotated, in steps of 3°, relative to the other (Fig. 3c, d, and Supplementary Figs 3e–g, 4 and 7). Maximal correlation was obtained for rotations ranging from 66° clockwise to 78° counterclockwise. With optimal rotation, the displacement of the population grid ranged from 5.7 to 27.5 cm (Fig. 3e, red vectors). The direction of the offset was uniformly distributed ($r = 0.60$, $P > 0.05$, Rayleigh test for randomness). On repeated trials in the same room (A versus A′), the peak of the cross-correlogram remained in the centre (range of displacements 0–2.3 cm; Fig. 3e, blue vectors). The shift between room A and room B, with optimal rotation, was significantly larger than that between repeated trials in the same room ($Z = 2.5$, $P < 0.01$). In five out of eight ensembles, the grid spacing was slightly different between rooms A and B (Supplementary Table 2); in those cases, the scaling was coherent between simultaneously recorded neurons (Supplementary Fig.

5c). Cross-correlograms for ensembles of CA3 cells in rooms A and B were flat (Fig. 3f and Supplementary Fig. 7). These observations suggest that grid fields moved and rotated, and sometimes expanded or contracted, when the rat was moved to a new location, but the intrinsic spatial phase structure of the grid cells was maintained. This differs from the orthogonalized representations of CA3 and indicates a possible pattern separation mechanism in the entorhinal inputs to the hippocampus.

We verified these results by examining whether ensemble activity in MEC was coherent in the temporal domain. Using only rats with simultaneous recordings from MEC and CA3, we compared the distribution of temporal population vectors, with population vectors defined as lists of cells that fired together across bins of 150 ms throughout the trial[13]. In MEC, the mean Shannon mutual information between trial and population vectors was no larger for comparisons of trials in different geometric shapes than for comparisons of
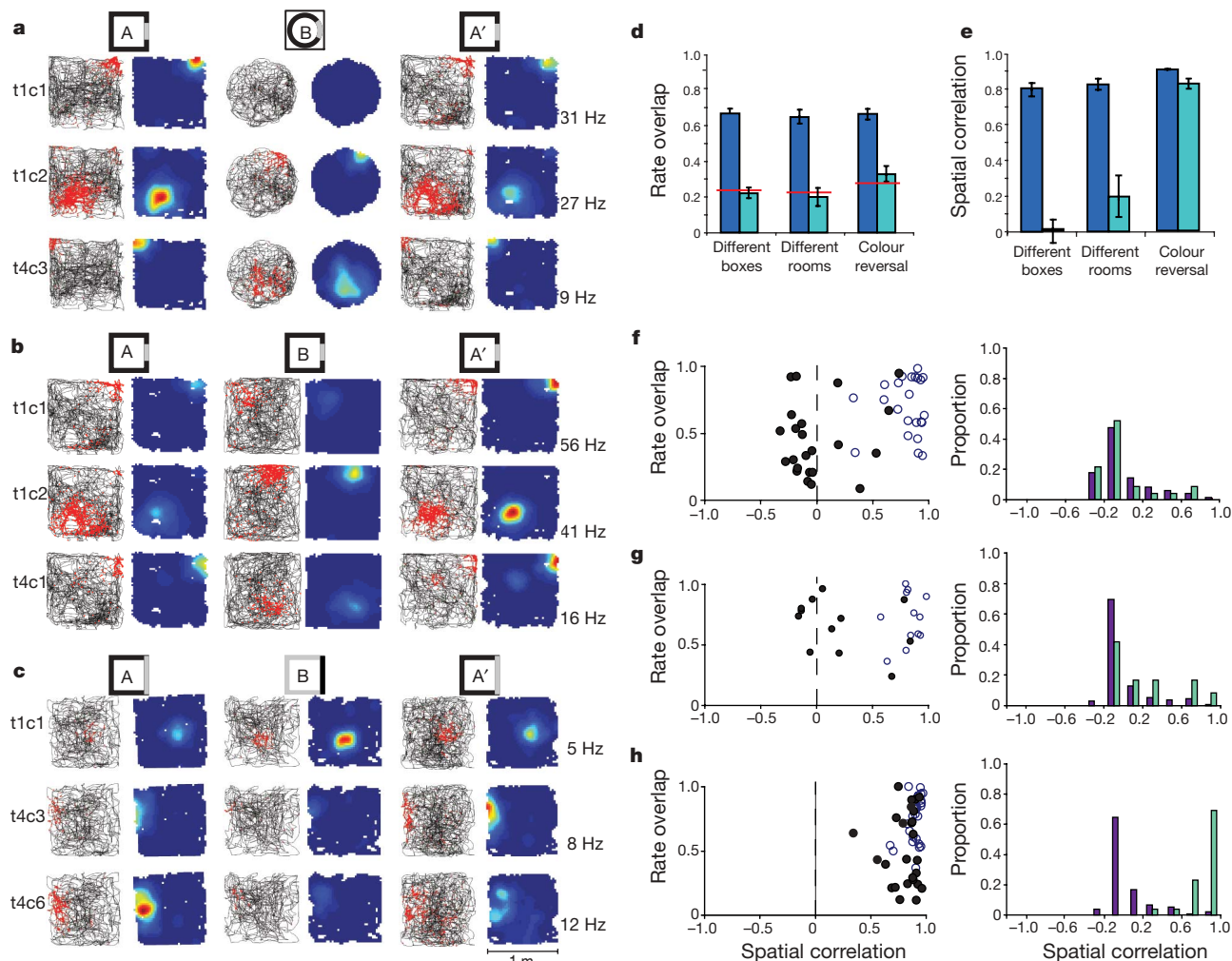
Figure 1 | Procedures for inducing global remapping and rate remapping in hippocampal area CA3. a–c, Firing fields of three simultaneously recorded CA3 cells during global remapping (different boxes in a constant location (a) or similar boxes in different rooms (b)) or rate remapping (different wall colours (c)). For complete cell samples see Supplementary Fig. 2a–f. Each row shows data from one active neuron (t, tetrode; c, cell). Each pair of columns shows, for each neuron, the location of spikes (red) on the trajectory of the rat (left) and a colour-coded map of the firing rate distribution (right; dark blue is zero, red is maximum; peak rate is indicated). **d,** Change in firing rates estimated by rate overlap (means ± s.e.m.). Red lines indicate chance levels. Blue bars, same environment (A–A′); green bars, different environments (A–B). **e,** Spatial correlation for cells with activity in both environments (means ± s.e.m.; best rotation). **f–h,** Rate overlap and spatial correlation in cells with activity in both environments. **f,** Different

boxes (as in **a**); **g,** different rooms (as in **b**); **h,** different colours (as in **c**). Scatter plots show rate overlap as a function of spatial correlation for individual CA3 cells (open circles, A versus A′; filled circles, A versus B). Frequency distributions show spatial correlations (green bars) compared with distributions for shuffled cell pairs (purple bars) from the same experiments. In the different-boxes task, 2 out of 23 spatial correlations exceeded the upper limit of the 95% confidence interval of the shuffled distribution. In the different-rooms task, 3 out of 12 correlations were outside the 95% confidence interval. In the rate remapping task, 25 out of 26 cells were outside the 95% confidence interval. The rate overlap between A and B is significantly higher in **f–h** than in **d** because the scatter plots in **f–h** do not include the large number of cells with low rates in one of the environments.

repeated trials in the same shape, and no larger for different rooms than for repeated trials in the same room (Supplementary Fig. 8; $t < 1.1$, $P > 0.05$, one-tailed paired t-tests), indicating that grid cells that are coactive in one environment may remain coactive in other environments too. In CA3, information about trial type was above 0.5 bits for square versus circle trials as well as trials in different rooms, significantly above the values obtained for repeated tests in the same environment (Supplementary Fig. 8; square versus circle: $t(7) = 5.6$, $P < 0.001$; two rooms: $t(4) = 3.8$, $P < 0.01$). The differentiation of coactivity patterns in CA3 further reinforces the view that incoming sensory input is decorrelated as it enters the hippocampus during global remapping.

In striking contrast with the coordinate shift during global remapping, grid vertices remained stable when place cells exhibited rate remapping, such as when the box colours were reversed. The peak of the cross-correlogram for trials with different colour configurations remained in the centre for six out of seven experiments (displacement 0–2.7 cm) and was slightly offset (7.1 cm) in one experiment (Fig. 4b, c, and Supplementary Figs 3h, i and 9). The peak displacement after colour reversal was not significantly different from the displacement observed in the same cells on trial pairs with identical colours (range 0–1.3 cm; $Z = 1.4$, $P > 0.05$). There was no change in the scale of the grid between 'black' and 'white' trials (Supplementary Table 2). There was also no significant change in the location of the cross-correlation peak in CA3 (median shift of 6.3 cm after colour reversal, in contrast with 4.9 cm with similar colours; $Z = 0.7$, not significant; Fig. 4d). Rate remapping in the hippocampus was not associated with significant rate changes

between the individual grid fields of the entorhinal neurons (Supplementary Fig. 10).

We finally examined whether global remapping in the hippocampus and grid realignment in the entorhinal cortex are coincident, because strong temporal contiguity would be expected if the phenomena were mechanistically related. Three rats, two of which had electrodes in both MEC and CA3, were trained in a square enclosure with lights on and off on alternating discrete sessions. The procedure induced global remapping between dark and light trials during the training phase in both rats with hippocampal electrodes (Supplementary Fig. 11a, b)[23]. The simultaneously recorded entorhinal representation was coherent but displaced and rotated between the trials. To evoke instantaneous global remapping, the room lights were turned on after 11 min of running in the dark condition on the test day. In one animal this caused a sudden reversion to the original hippocampal map associated with the light condition. This instance of global hippocampal remapping coincided with an equally fast realignment of the entorhinal map, expressed as a loss of spatial correlation between grid fields recorded before and after the light was turned on (Supplementary Fig. 11a, c). In the second rat, all hippocampal and entorhinal neurons maintained their firing fields when the darkness was terminated[23], presumably reflecting a continued influence of self-motion information on the location of firing in grid cells and place cells[16]. To reset the path integrator, the rat was lifted up on a pedestal for 1 min. The temporary disruption of running caused an immediate reversion to the original hippocampal map, associated with the light condition, when the rat was returned to the box. At the same time, a coherent shift and rotation of the population grid was
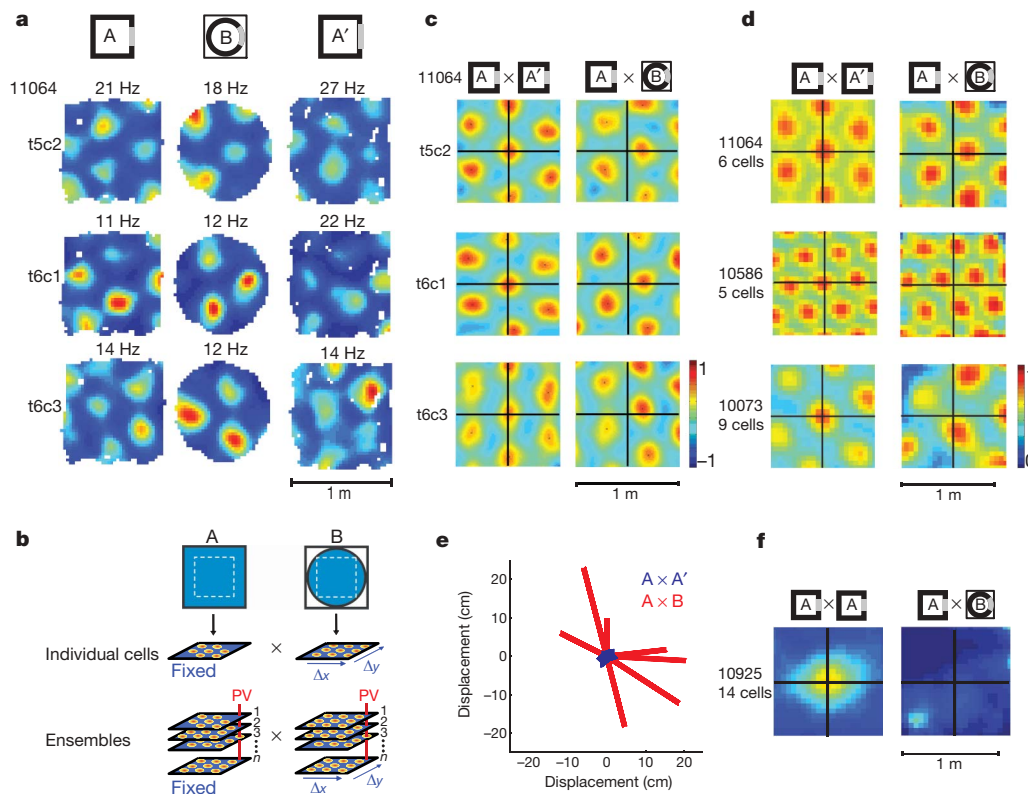
Figure 2 | Realignment of entorhinal grid fields during hippocampal global remapping between different boxes in the same location. a, Colour-coded firing rate maps for three representative simultaneously recorded MEC cells (a complete cell sample is given in Supplementary Fig. 3a). Five-digit numbers (for example 11064) refer to rat identities. **b**, Schematic illustration of the procedure for cross-correlation of rate maps based on individual rate maps (top) or stacks of rate maps (bottom). Cross-correlation matrices were determined by shifting the two maps relative to one another in steps of 5 cm in both the x and y directions within the common stippled area. PV,

population vector (defined for each 5×5 cm^2 bin). **c, d**, Colour-coded cross-correlation matrices for representative individual cells (**c**; same cells as in **a**) and cell ensembles (**d**) during repeated testing in the same box (A versus A′) or in different boxes (A versus B) (for all other experiments see Supplementary Fig. 3c). **e**, Vector diagram showing distribution across experiments of the direction and the distance of the grid displacement, measured from the origin to the nearest peak of the cross-correlogram. **f**, Representative cross-correlogram for CA3 ensembles.
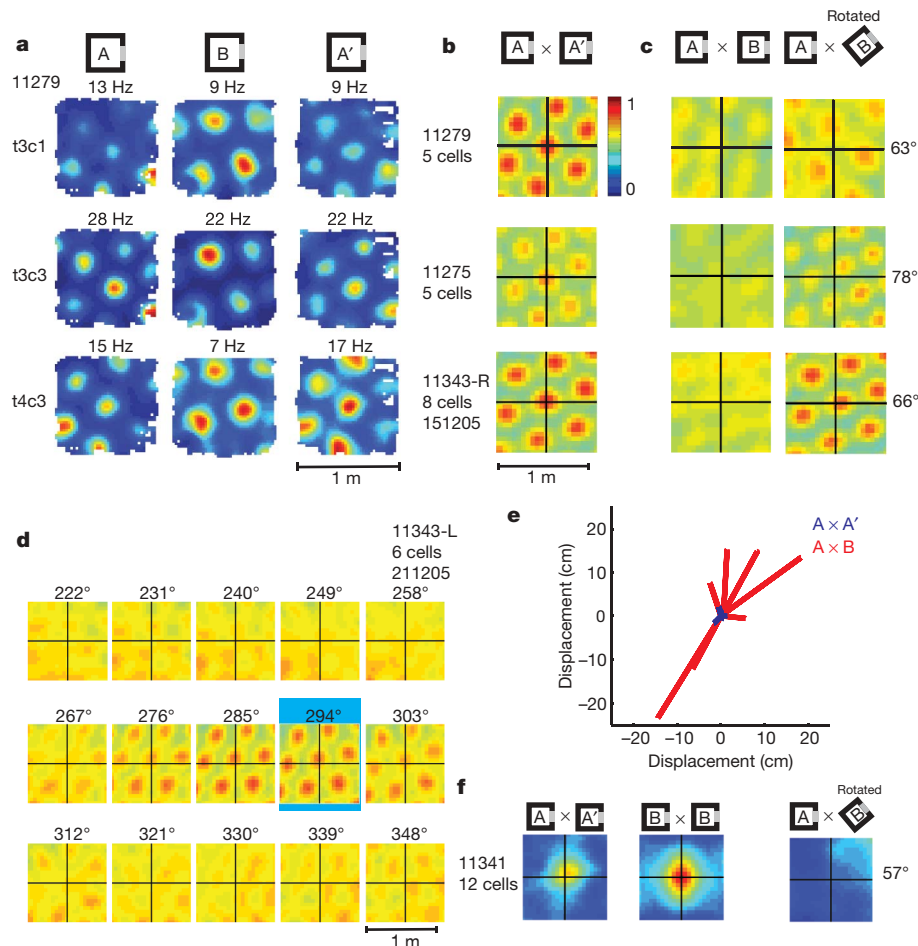
Figure 3 | Realignment of entorhinal grid fields during hippocampal global remapping between two rooms. a, Rate maps for representative simultaneously recorded MEC cells (a complete cell sample is given in Supplementary Fig. 3e). **b, c,** Cross-correlation matrices for representative grid cell ensembles. **b,** Repeated trials in room A (A–A′); **c,** room A versus room B. The right column in **c** shows cross-correlograms for the rotation that gave maximal grid structure (left column, no rotation). For all other experiments see Supplementary Fig. 3f, g. **d,** Cross-correlogram for incremental amounts of rotation in one experiment (blue, maximal correlation). **e,** Vector diagram showing distribution across experiments of direction and distance of grid displacement after optimal rotation. **f,** Representative cross-correlograms for simultaneously recorded CA3 pyramidal cells.
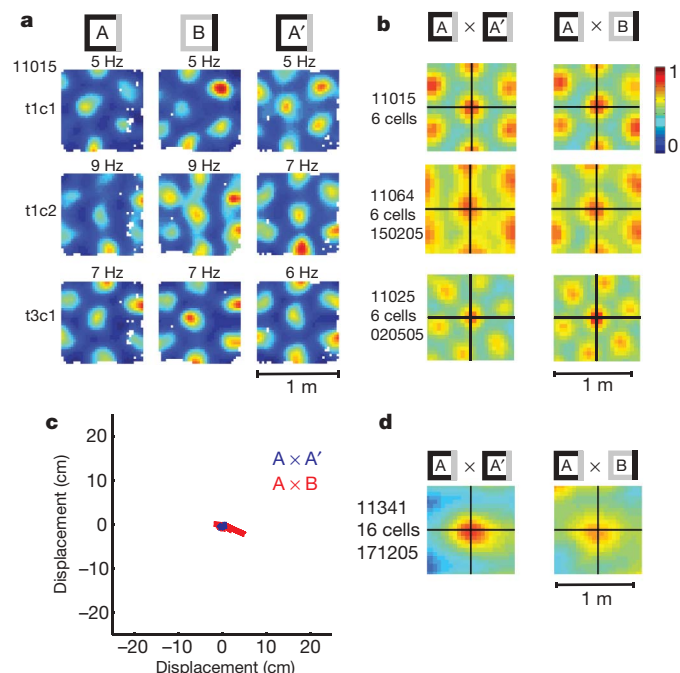
observed in MEC. The coincidence of the changes in entorhinal and hippocampal population dynamics in the two animals indicates that the changes may be part of a single integrated process.

First, these results indicate that local ensembles of grid cells may have a rigid spatial phase relationship. The repetition of the same sequence of active units as an animal follows a similar route in different arenas supports the hypothesis that the grid network is a universal metric for path-integration-based navigation[16,19,24–26]. Second, the association between hippocampal global remapping and entorhinal grid realignment implies network dynamics in the MEC as a possible determinant of orthogonalization processes in hippocampal place cells (Supplementary Fig. 12). The stability of entorhinal grid cells during rate remapping in the hippocampus raises the possibility that the rate orthogonalization originates in the dentate gyrus and CA3 (ref. 27), perhaps as a consequence of convergence between

Figure 4 | Unaltered alignment of entorhinal grid fields during rate remapping in CA3. a, Rate maps for representative simultaneously recorded grid cells in the different-colours task (a complete cell sample is given in Supplementary Fig. 3h). **b,** Cross-correlation maps for representative grid cell ensembles (for all other experiments see Supplementary Fig. 3i). Left column, trials with similar wall colours (A–A′); right, trials with different wall colours (A–B). **c,** Vector diagram showing distribution across experiments of direction and distance of grid displacement. **d,** Representative cross-correlogram for CA3 ensembles.

spatial signals from grid cells in MEC and nonspatial signals from neurons in the lateral entorhinal cortex[28,29]. The possible recoding of spatial information from a single, universal neural representation in the entorhinal cortex onto statistically independent context-sensitive cell ensembles in high-capacity networks of the hippocampus[30] is probably crucial for the successful storage of episodic memory (see Supplementary Discussion).

METHODS

Neuronal ensemble activity was recorded with tetrodes in 19 Long Evans rats implanted with microdrives in dorsocaudal MEC and/or dorsal CA3 of the hippocampus. Spike-triggered activity was sampled in blocks of 10 or 20 min while rats chased soft food crumbs in familiar enclosures of different shapes, textures and colours. Global remapping was induced by alternating between square and circular recording enclosures at a constant location within the same room, by testing the rat in similar square boxes in two rooms with distinct background cues[13], or by training the rat on separate trials in the same environment in light and darkness[23]. Rate remapping was induced by training the animals in a square enclosure with exchangeable walls of different colour[14]. Detailed experimental and analytical procedures are provided in Supplementary Methods.

1. Marr, D. Simple memory: a theory of archicortex. *Phil. Trans. R. Soc. Lond. B* **262**, 23–81 (1971).
2. McNaughton, B. L. & Nadel, L. in *Neuroscience and Connectionist Theory* (eds Gluck, M. A. & Rumelhart, D. E.) 1–63 (Erlbaum, Hillsdale, NJ, 1989).
3. Treves, A. & Rolls, E. T. Computational constraints suggest the need for two distinct input systems to the hippocampal CA3 network. *Hippocampus* **2**, 189–199 (1992).
4. Kesner, R. P., Gilbert, P. E. & Wallenstein, G. V. Testing neural network models of memory with behavioral experiments. *Curr. Opin. Neurobiol.* **10**, 260–265 (2000).
5. Leutgeb, S., Leutgeb, J. K., Moser, M.-B. & Moser, E. I. Place cells, spatial maps and the population code for memory. *Curr. Opin. Neurobiol.* **15**, 738–746 (2005).
6. O'Keefe, J. & Nadel, L. *The Hippocampus as a Cognitive Map* (Clarendon, Oxford, 1978).
7. Bostock, E., Muller, R. U. & Kubie, J. L. Experience-dependent modifications of hippocampal place cell firing. *Hippocampus* **1**, 193–205 (1991).
8. Markus, E. J. *et al.* Interactions between location and task affect the spatial and directional firing of hippocampal neurons. *J. Neurosci.* **15**, 7079–7094 (1995).
9. Kentros, C. *et al.* Abolition of long-term stability of new hippocampal place cell maps by NMDA receptor blockade. *Science* **280**, 2121–2126 (1998).
10. Wood, E. R., Dudchenko, P. A., Robitsek, R. J. & Eichenbaum, H. Hippocampal neurons encode information about different types of memory episodes occurring in the same location. *Neuron* **27**, 623–633 (2000).
11. Frank, L. M., Brown, E. N. & Wilson, M. Trajectory encoding in the hippocampus and entorhinal cortex. *Neuron* **27**, 169–178 (2000).
12. Lever, C., Wills, T., Cacucci, F., Burgess, N. & O'Keefe, J. Long-term plasticity in hippocampal place-cell representation of environmental geometry. *Nature* **416**, 90–94 (2002).
13. Leutgeb, S., Leutgeb, J. K., Treves, A., Moser, M.-B. & Moser, E. I. Distinct ensemble codes in hippocampal areas CA3 and CA1. *Science* **305**, 1295–1298 (2004).
14. Leutgeb, S. *et al.* Independent codes for spatial and episodic memory in the hippocampus. *Science* **309**, 619–623 (2005).
15. Fyhn, M., Molden, S., Witter, M. P., Moser, E. I. & Moser, M.-B. Spatial representation in the entorhinal cortex. *Science* **305**, 1258–1264 (2004).
16. Hafting, T., Fyhn, M., Molden, S., Moser, M.-B. & Moser, E. I. Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
17. Steffenach, H.-A., Witter, M. P., Moser, M.-B. & Moser, E. I. Spatial memory in the rat requires the dorsolateral band of the entorhinal cortex. *Neuron* **45**, 301–313 (2005).
18. Sargolini, F. *et al.* Conjunctive representation of position, direction and velocity in entorhinal cortex. *Science* **312**, 754–758 (2006).
19. McNaughton, B. L., Battaglia, F. P., Jensen, O., Moser, E. I. & Moser, M.-B. Path integration and the neural basis of the 'cognitive map'. *Nature Rev. Neurosci.* **7**, 663–678 (2006).
20. Solstad, T., Moser, E. I. & Einevoll, G. T. From grid cells to place cells: a mathematical model. *Hippocampus* **16**, 1026–1031 (2006).
21. Wills, T. J., Lever, C., Cacucci, F., Burgess, N. & O'Keefe, J. Attractor dynamics in the hippocampal representation of the local environment. *Science* **308**, 873–876 (2005).
22. Anderson, M. I. & Jeffery, K. J. Heterogeneous modulation of place cell firing by changes in context. *J. Neurosci.* **23**, 8827–8835 (2003).
23. Quirk, G. J., Muller, R. U. & Kubie, J. L. The firing of hippocampal place cells in the dark depends on the rat's recent experience. *J. Neurosci.* **10**, 2008–2017 (1990).
24. Redish, A. D. & Touretzky, D. S. Cognitive maps beyond the hippocampus. *Hippocampus* **7**, 15–35 (1997).
25. Sharp, P. E. Complimentary roles for hippocampal versus subicular/entorhinal place cells in coding place, context, and events. *Hippocampus* **9**, 432–443 (1999).
26. Fuhs, M. C. & Touretzky, D. S. A spin glass model of path integration in rat medial entorhinal cortex. *J. Neurosci.* **26**, 4266–4276 (2006).
27. Leutgeb, J. K., Leutgeb, S., Moser, M.-B. & Moser, E. I. Pattern separation in dentate gyrus and CA3 of the hippocampus. *Science* **315**, doi:10.1126/science.1135801 (16 February 2007)
28. Hargreaves, E. L., Rao, G., Lee, I. & Knierim, J. J. Major dissociation between medial and lateral entorhinal input to the dorsal hippocampus. *Science* **308**, 1792–1794 (2005).
29. Naber, P. A., Caballero-Bleda, M., Jorritsma-Byham, B. & Witter, M. P. Parallel input to the hippocampal memory system through peri- and postrhinal cortices. *Neuroreport* **8**, 2617–2621 (1997).
30. Battaglia, F. P. & Treves, A. Attractor neural networks storing multiple space representations: a model for hippocampal place fields. *Phys. Rev. E Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.* **58**, 7738–7753 (1998).

LETTERS

A cytosolic *trans*-activation domain essential for ammonium uptake

D. Loqué[1]*, S. Lalonde[1]*, L. L. Looger[1]*, N. von Wirén[2] & W. B. Frommer[1]

Polytopic membrane proteins are essential for cellular uptake and release of nutrients. To prevent toxic accumulation, rapid shut-off mechanisms are required. Here we show that the soluble cytosolic carboxy terminus of an oligomeric ammonium transporter from *Arabidopsis thaliana* serves as an allosteric regulator essential for function; mutations in the C-terminal domain, conserved between bacteria, fungi and plants, led to loss of transport activity. When co-expressed with intact transporters, mutants inactivated functional subunits, but left their stability unaffected. Co-expression of two inactive transporters, one with a defective pore, the other with an ablated C terminus, reconstituted activity. The crystal structure of an *Archaeoglobus fulgidus* ammonium transporter (AMT)[1] suggests that the C terminus interacts physically with cytosolic loops of the neighbouring subunit. Phosphorylation of conserved sites in the C terminus[2] are proposed as the cognate control mechanism. Conformational coupling between monomers provides a mechanism for tight regulation, for increasing the dynamic range of sensing and memorizing prior events, and may be a general mechanism for transporter regulation.

Some transporters oligomerize to form pores, whereas in many metabolite transporters (including AMT ammonium transporters[3–5]) each subunit in the oligomer forms a functional pore[6–8]. Oligomerization is commonly attributed to the hydrophobic effect driving association in lipid bilayers[9]. The unexpected kinetic behaviour of mammalian glucose transporters, originally described as uniporters, suggested that the quaternary state of homomeric transporters may affect function by yet unknown mechanisms[10]. Could allostery between subunits explain the prevalence of oligomers in membranes and the unconventional behaviour of some transporters[6]? We began from the observation that a mutation in the cytosolic C terminus (CCT) of an oligomeric AMT led to inactivation and epistatic suppression of a co-expressed isoform[4,11]. Two hypotheses could explain epistasy: (1) inter-subunit allostery, or (2) defective secretion preventing the complex from reaching the plasma membrane. Here we provide evidence that the CCT of the *Arabidopsis* ammonium transporter AMT1;1 acts as an allosteric switch between neighbouring monomers (Supplementary Fig. 1). We also show that modification of the CCT, presumably by phosphorylation, provides a means of regulating activity. Allosteric *trans*-activation (ATA) allows rapid inactivation of a multi-pore transporter complex by a single regulatory event, and for memory, as multiple regulatory states can be stored in the cache of the symmetric oligomer[12]. This mechanism allows rapid transporter shut-off to protect against over-accumulation or membrane depolarization at high external ammonium.

Members of the AMT/MEP/Rh superfamily, which sense and transport ammoniacal nitrogen in all organisms, oligomerize[4,13,14]. Previous investigations suggested inter-monomer cross-talk in yeast and plant family members without providing a mechanism: a Gly413Asp mutation in the CCT inactivated the yeast homologue MEP1 and led to epistatic suppression of its paralogue MEP3 (ref. 11). Equivalent mutants of *Arabidopsis* (Gly456Asp) and tomato[4] homologues were non-functional and *trans*-inactivated wild-type transporters (Fig. 1a; Supplementary Information). Stability and targeting of AMT1;1 seemed unaffected (Fig. 1b), favouring the ATA hypothesis. Compared with the extracellular loops, the first 24–30 amino acids of the CCT and the intracellular loops are more highly conserved among bacteria, plants, and fungi (Supplementary Fig. 2). AMT1;1 variants mutated in the CCT were expressed alone or with wild type (Fig. 2a, b). Truncation mutants
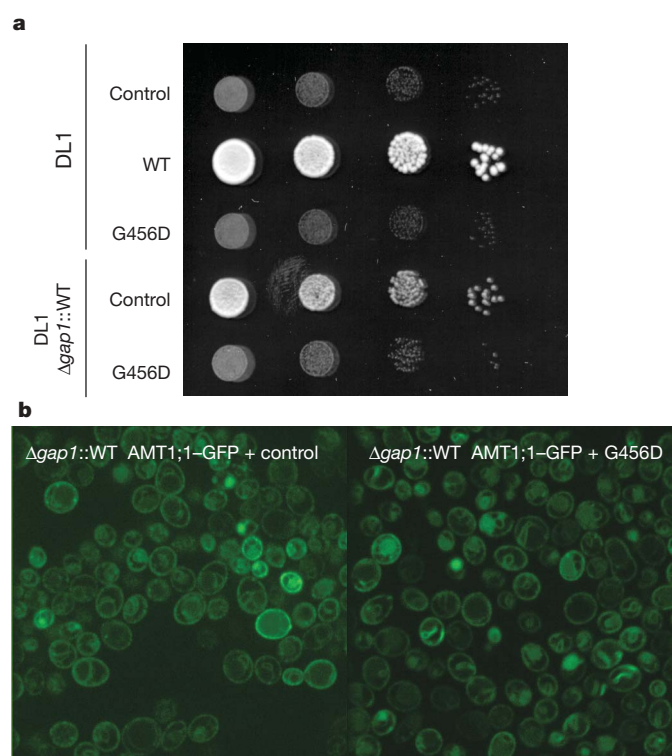
Figure 1 | *Arabidopsis* AMT1;1 mutant functionality measured by their ability to confer growth to DL1 (Δ*gap1* Δ*mep1-3*) on 2 mM ammonium as sole N-source. a, Complementation by *pDR196*-expressed G456D of DL1 and an integrated wild-type strain (DL1 Δ*gap1*::WT). DL1 was transformed with *Arabidopsis* AMT1;1 mutants (episomal *pDRf1-GW*); incubated on 2 mM ammonium for 4 d; see Supplementary Fig. 15. Control, empty vector; WT, wild type *Arabidopsis* AMT1;1. **b**, Confocal section of yeast with an integrated *Arabidopsis* AMT1;1–GFP fusion (Δ*gap1*::WT AMT1;1–GFP) co-expressing *Arabidopsis* AMT1;1 mutants (episomal *pDRf1-GW*).

[1]Carnegie Institution, 260 Panama St, Stanford, California 94305, USA. [2]Institute for Plant Nutrition, University of Hohenheim, Stuttgart 70593, Germany.
*These authors contributed equally to this work.

retaining the last amino acid (Tyr 469) of the conserved CCT were functional; truncation before Tyr 469 led to profound loss of activity (Fig. 2a). Further truncation led to *trans*-inactivation of wild-type AMT1;1 (Fig. 2b); similarly, CCT insertions abolished activity (Supplementary Fig. 3). In the same way, insertions into loops L3–4 and L5–6 blocked activity (Supplementary Fig. 4). This contrasts with lactose permease[15] and sucrose transporters[16,17], where loop mutations are well tolerated. Structural evidence supports an ATA interaction: the CCT of *Archaeoglobus* Amt-1 is well structured[1], forming main- and side-chain interactions with L1–2, L3–4, and L5–6 cytoplasmic loops of its own monomer, as well as with L1–2, L5–6, and L7–8 loops of the adjacent monomer[1]. We predict that the intra- and inter-subunit interactions are important for allosteric linkage. A homology model of *Arabidopsis* AMT1;1 based on the *Archaeoglobus* Amt-1 structure (Fig. 3; Supplementary Figs 5–7) suggests that the overall fold is similar with structurally conserved transmembrane spans (TMSs), cytoplasmic loops and CCT, whereas extracellular loops are more elaborated (Supplementary Fig. 5). Mutagenesis data support the modelled CCT/loop interactions. The critical CCT residue Tyr 469 is modelled to interact with the adjacent L5–6 loop, and Tyr 467 (CCT) is predicted to form a hydrogen bond with His 239 (L5–6) within each monomer (Fig. 3b; Supplementary Figs 5–7). Conservative mutation of Tyr 467 or His 239 to phenylalanine yielded non-functional and *trans*-inactivating transporters (Fig. 2c). A targeted suppressor screen of the defective Tyr467Phe transporter led only to His239Cys (Supplementary Fig. 8). *Arabidopsis* AMT1;1(His239Cys) was non-functional; the reverse screen for His239Cys-suppressors identified only Tyr467Phe

(Supplementary Fig. 8). Thus, precise positioning of the C terminus, whether by hydrogen bonding or non-polar interactions, seems critical for functionality.

Quantitative comparison of the uptake rates of integrated and episomal wild-type *Arabidopsis* AMT1;1 allowed estimation of the relative copy number and thus the population distribution in co-expressions (Supplementary Fig. 9). This analysis indicates that in the case of *trans*-inactivation of an integrated wild-type transporter by an episomal mutant, mutant$_2$wt$_1$ and mutant$_1$wt$_2$ hetero-trimers seem to be inactive (Supplementary Information). Apparent non-functionality of the wild-type monomer in a complex with one or two
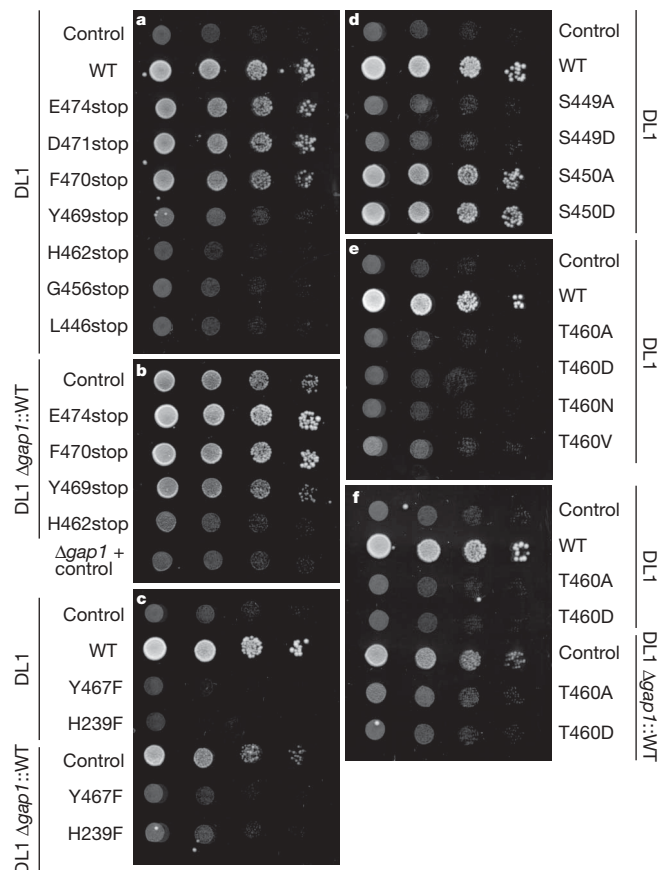
Figure 2 | Functional characterization of AMT1;1 mutants.
a–e, Complementation by (**a**) CCT truncation mutants (designated X000stop), (**b**) truncation mutants co-expressed with the integrated wild type, (**c**) H239F and Y467F mutants expressed alone or with wild type, (**d**) S449 and S450 mutants, (**e**) T460 mutants expressed alone, or (**f**) co-expression of T460A and T460D with wild type. For conditions see Fig. 1.
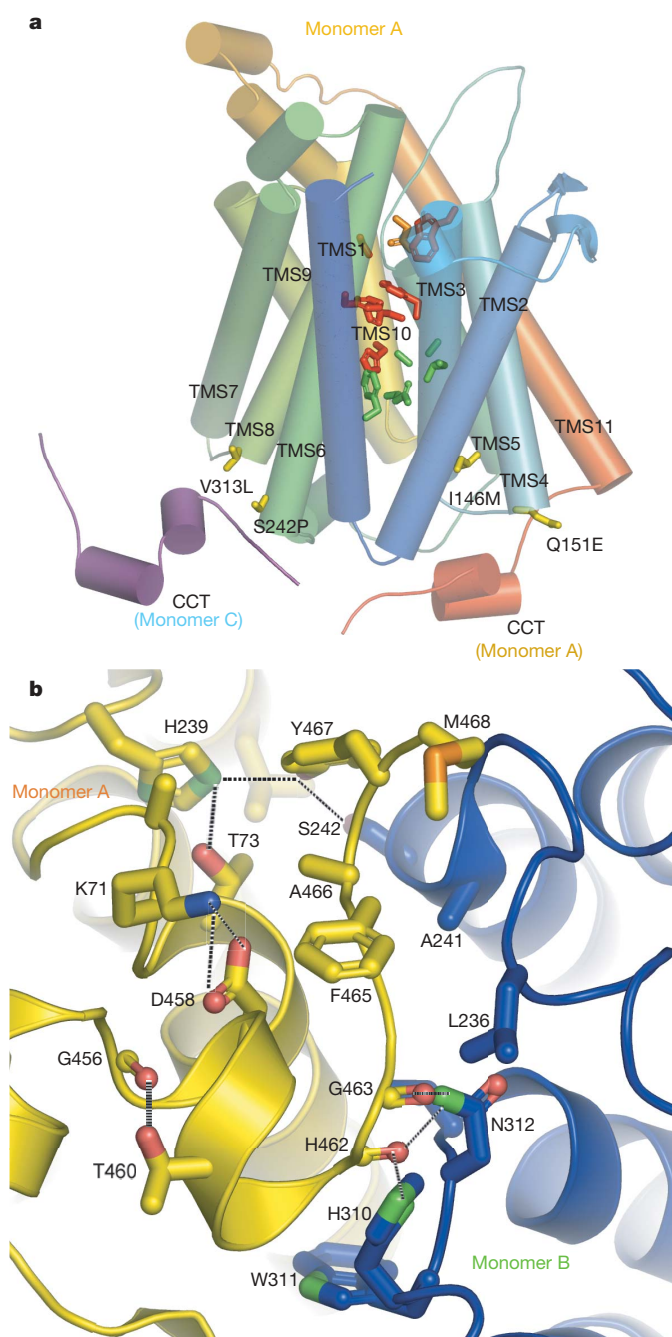
Figure 3 | Structural model of *Arabidopsis* AMT1;1. a, Monomer with adjacent CCT (substrate channel amino acids, red; permissive amino acids, green; CCT-stabilizing amino acids, yellow; D198, orange; adjacent CCT, purple). **b**, Proposed interactions of the CCT, within and between monomers. Monomer A, yellow; Monomer B, blue; hydrogen bonds, dashed lines.

mutants supports tight allosteric coupling in the trimer. The failure of defective-terminus monomers to be *trans*-activated by wild-type CCTs further supports coupling between activation states of individual subunits.

Rapid inactivation of uptake after ammonium addition suggests post-translational regulation of ammonium transport activity[18–20]. The CCT contains three potential phospho-regulatory Ser/Thr residues. Mutation of Ser 449 (highly conserved; predicted to hydrogen bond within the CCT) to aspartate or alanine inhibited transport activity, whereas Ser 450 (poorly conserved, solvent-exposed) mutants retained activity (Fig. 2d). *Arabidopsis* AMT1;1 contains an unconventional kinase recognition site at Thr 460 (predicted to hydrogen bond within the CCT; Fig. 3b). Mutation of Thr 460 to Ala, Asn or Val reduced transport activity (Fig. 2e, Supplementary Fig. 10). Mutation to Asp (phosphorylation mimic) yielded non-functional and *trans*-inactivating transporters (Fig. 2e, f). Abundance of Thr460Ala and Thr460Asp mutant proteins was similar to wild type, and plasma membrane targeting of a functional green fluorescence protein (GFP)-tagged wild-type transporter was unaffected by co-expressed Thr460Ala or Thr460Asp (Supplementary Fig. 11). *In-vivo* phosphorylation of Thr 460 in *Arabidopsis* cell cultures grown in high ammonium concentrations[2] implicates it as a cognate phospho-regulatory site. Wild-type AMTs are functional in heterologous systems, presumably lacking AMT1;1-specific kinases[4,21,22]; thus the default regulatory state of the CCT is predicted to be activating. Phosphorylation may inactivate *Arabidopsis* AMT1;1 nonlinearly as part of a feedback-inhibition loop. Not all AMT/MEP proteins retain the same kinase recognition site—others may employ phosphorylation sites upstream (for example, positions corresponding to Ser 449 in *Arabidopsis* AMT1;1) or in the loops. Alternatively, bacterial AMTs are down-regulated by interaction with trimeric GlnK/P_{II} proteins[23], presumably forcing the complex into an inactive state.

To discover mutations compensating for a defective CCT, a multi-copy suppressor screen was performed using the inactive Thr460Ala transporter (Supplementary Information). Thirty-three suppressors were discovered (15 mutations at 9 positions; Supplementary Table 1, Supplementary Fig. 12); suppressors clustered in two domains: at the cytosolic interface of TMSs 8–9 and at the centre of TMSs 1–4 (Fig. 3a). No revertants were identified, which is consistent with inactivity of $(Thr460Ala)_2wt_1$ and $(Thr460Ala)_1wt_2$ complexes. All suppressor mutations were introduced into the Thr460Asp background. All $(Thr460Ala/suppressor)_3$ transporters and five $(Thr460Asp/suppressor)_3$ trimers, were functional (Supplementary Fig. 12). These five mutations cluster in three consecutive positions in TMS 3 and two nearby positions in TMS 1, close to the ammonium-conducting and hydrophobic constriction side chains[1] (Fig. 3a). Three out of these five mutations (Phe60Ser, Ile136Phe and Ala137Asp) were also found in a suppressor screen for restoration of activity of the truncated AMT1;1(Gly456stop) (Supplementary Fig. 13), suggesting perturbation of the immediate channel environment and creation of a constitutive state devoid of allosteric modulation. The Thr460Ala suppressors Ile146Met and Gln151Glu seem to stabilize the CCT through intra-monomer contacts, whereas Ser242Pro and Val313Leu modulate interaction with the adjacent CCT (Fig. 3a). We propose that these four mutations, which recover activity of Thr460Ala but not Thr460Asp, restore CCT-to-loop interactions, increasing the propensity of a subunit either to *trans*-activate or to be *trans*-activated (for example, by stabilizing L5–6). The loss-of-CCT-interaction by Thr460Ala mutation may be restored by compensatory changes, but the Thr460Asp mutation introduces a charge mutation to the CCT that may not be rescued.

To separate the inherent transport capacity of an individual monomer from the *trans*-activating effects of its C terminus, we tried to reconstitute a functional complex from a *trans*-activation-deficient mutant (Thr460Ala) and a transport-deficient version of the *trans*-activation-suppressor Ser242Pro. A conductance-compromised variant of *Arabidopsis* AMT1;1 was created by mutagenesis of highly conserved Asp 198 to asparagine. Asp 198 makes two stabilizing hydrogen

bonds to TMS 5 and has been proposed as an ammonium-recruitment site[23,24] (Fig. 3a). *Arabidopsis* AMT1;1(Asp198Asn) was inactive (Fig. 4a; Supplementary Fig. 14); equivalent mutations in yeast MEPs, *Escherichia coli* AmtB, and human RhAG produced inactive but stable proteins[23,24]. Co-expression of the inactive Asp198Asn and Thr460Ala mutants created a non-functional transporter (Fig. 4b, Supplementary Fig. 14), further demonstrating that a wild-type C terminus (as in Asp198Asn) is insufficient for *trans*-activation when in a complex with *trans*-activation-deficient monomers (Thr460Ala). The non-functional Ser242Pro/Asp198Asn double mutant, when co-expressed with the defective Thr460Ala, reconstituted a functional complex (Fig. 4b). The CCT-stabilizing mutation Ser242Pro seems to compensate for the destabilizing mutation Thr460Ala in a trimer, driving the CCT back into a *trans*-activating conformation.

Altogether, we propose that intimate contacts between CCT, L5–6 and adjacent loops, both in its own monomer and the neighbour, serve as allosteric switches linking the conformational states of the monomers in the complex, allowing a concerted (Monod–Wyman–Changeux) conformational change (Supplementary Fig. 1)[25]. The interactions of the CCT may modulate the position of the ammonium-conducting or hydrophobic constriction side chains[1,5], controlling substrate access to, and/or transport through, the pore.

This super-linear regulatory mechanism may have evolved to allow for rapid responses to regulatory cues, propagating conformational changes throughout a complex following a single signalling event. This mechanism, judged from its conservation, has been preserved from cyanobacteria to higher plants. Rapid inactivation of transport in potentially toxic environments may be the common function. Identical phosphorylation sites in each monomer provide for memory of previous exposure to ammonium, and if coupled to signalling, may lead to oscillatory behaviour[12]. Because ammonium transporters have been implicated in signalling, allosteric control may increase the dynamic range for signal transduction[23]. Apart from the interesting regulatory aspects, the findings have implications for the interpretation of AMT structures regarding conformational dynamics and transport mechanism[5].

ATA may have implications beyond nitrogen transport. The C-terminal domain of PIP2 aquaporins interacts with the adjacent monomer, and phosphorylation of Ser 274 activates transport[26].
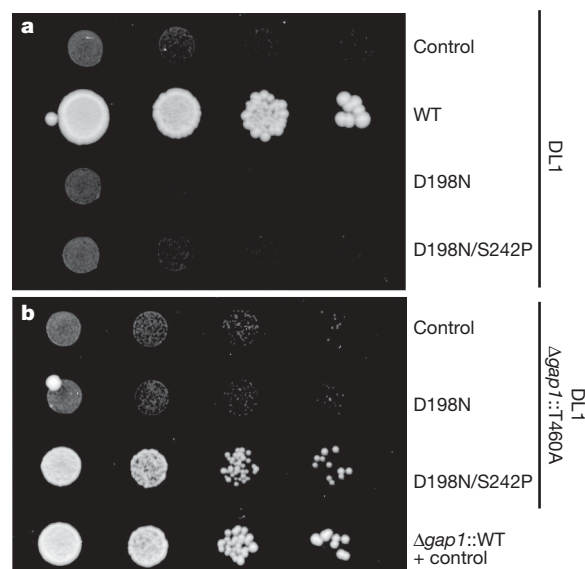

Figure 4 | Co-expression of the *trans*-activation-deficient T460A and ammonium-recruitment-deficient D198N mutants reconstitutes a functional complex. D198N and D198N/S242P were expressed in DL1 (**a**) or DL1 Δ*gap1*::T460A (**b**). Growth on 2 mM ammonium was assayed after 6 d (see Supplementary Fig. 16). Control, *pDRf1*; WT, wild-type *Arabidopsis* AMT1;1.

Many membrane proteins function as oligomers; ATA may provide an explanation for the switch in the transport mechanism of *Kluyveromyces lactis* lactose permease[6]. Auto-regulatory domains are found in H^+- and Ca^{2+}-ATPases[27,28] and CAX[29]; such extensions may function in *trans*, similar to the oligomeric mammalian calmodulin kinase[12], rather than as *sensu stricto* auto-regulatory domains. As for mammalian and plant calcium calmodulin kinases, *trans*-regulatory domains may decode oscillatory signals and provide for memory[12,30].

METHODS

Yeast expression and coexpression. Yeast strain 31019b (*Δmep1 mep2::LEU2 mep3::KanMX2 ura3*)[11] was transformed with wild-type or mutant AMT1;1 using high copy episomal or single copy integrative expression vectors. To test for functionality, cells were grown in liquid YNB medium supplemented with 3% glucose and 1 mM arginine for one day, diluted $\times 10^1$–10^4 and dropped on solid YNB medium supplemented with 0.2, 0.4 or 2 mM ammonium chloride or 1 mM arginine. Cells were incubated 3–6 days at 28 °C. Targeted or non-targeted suppressors of defective AMT1;1 mutants were identified under selective conditions.

Protein analyses. Yeast cells expressing the AMT1;1 variants were harvested by centrifugation at $5,000g$ for 10 min at 4 °C, washed and resuspended in 20 mM Tris-HCl, 10 mM $MgCl_2$, 1 mM EDTA, 5% Glycerol, 1 mM DTT, 1 mM PMSF and $1\times$ protease inhibitor cocktail (Sigma-Aldrich), pH 8 and disrupted with glass beads. Microsomal fractions were harvested at $100,000g$. The sediment was resuspended in 20 mM Tris-Cl, 0.1 mM EDTA, 10% glycerol, 100 mM KCl, 1 mM DTT, 1 mM PMSF and $1\times$ protease inhibitor cocktail, pH 7.5. Protein gel blots were used to detect the protein using a polyclonal antiserum raised against a C-terminal AMT1;1 peptide (N-RRVEPRSPSPSGANT-C).

Confocal microscopy. The subcellular distribution of an AMT1;1–GFP fusion was analysed using a spinning disk confocal microscope.

1. Andrade, S. L., Dickmanns, A., Ficner, R. & Einsle, O. Crystal structure of the archaeal ammonium transporter Amt-1 from *Archaeoglobus fulgidus. Proc. Natl Acad. Sci. USA* **102**, 14994–14999 (2005).
2. Nühse, T. S., Stensballe, A., Jensen, O. N. & Peck, S. C. Phosphoproteomics of the *Arabidopsis* plasma membrane and a new phosphorylation site database. *Plant Cell* **16**, 2394–2405 (2004).
3. Blakey, D. *et al.* Purification of the *Escherichia coli* ammonium transporter AmtB reveals a trimeric stoichiometry. *Biochem. J.* **364**, 527–535 (2002).
4. Ludewig, U. *et al.* Homo- and hetero-oligomerization of ammonium transporter-1 NH_4^+ uniporters. *J. Biol. Chem.* **278**, 45603–45610 (2003).
5. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science* **305**, 1587–1594 (2004).
6. Veenhoff, L. M., Heuberger, E. H. & Poolman, B. Quaternary structure and function of transport proteins. *Trends Biochem. Sci.* **27**, 242–249 (2002).
7. Jiang, Y. *et al.* X-ray structure of a voltage-dependent K^+ channel. *Nature* **423**, 33–41 (2003).
8. Borgnia, M., Nielsen, S., Engel, A. & Agre, P. Cellular and molecular biology of the aquaporin water channels. *Annu. Rev. Biochem.* **68**, 425–458 (1999).
9. Grasberger, B., Minton, A. P., DeLisi, C. & Metzger, H. Interaction between proteins localized in membranes. *Proc. Natl Acad. Sci. USA* **83**, 6258–6262 (1986).
10. Hamill, S., Cloherty, E. K. & Carruthers, A. The human erythrocyte sugar transporter presents two sugar import sites. *Biochemistry* **38**, 16974–16983 (1999).
11. Marini, A. M., Springael, J. Y., Frommer, W. B. & André, B. Cross-talk between ammonium transporters in yeast and interference by the soybean SAT1 protein. *Mol. Microbiol.* **35**, 378–385 (2000).
12. Schulman, H., Hanson, P. I. & Meyer, T. Decoding calcium signals by multifunctional CaM kinase. *Cell Calcium* **13**, 401–411 (1992).
13. Coutts, G., Thomas, G., Blakey, D. & Merrick, M. Membrane sequestration of the signal transduction protein GlnK by the ammonium transporter AmtB. *EMBO J.* **21**, 536–545 (2002).
14. Conroy, M. J. *et al.* Electron and atomic force microscopy of the trimeric ammonium transporter AmtB. *EMBO Rep.* **5**, 1153–1158 (2004).
15. Zen, K. H., McKenna, E., Bibi, E., Hardy, D. & Kaback, H. R. Expression of lactose permease in contiguous fragments as a probe for membrane-spanning domains. *Biochemistry* **33**, 8198–8206 (1994).
16. Reinders, A. *et al.* Intra- and intermolecular interactions in sucrose transporters at the plasma membrane detected by the split-ubiquitin system and functional assays. *Structure* **10**, 763–772 (2002).
17. Schulze, W. X., Reinders, A., Ward, J., Lalonde, S. & Frommer, W. B. Interactions between co-expressed *Arabidopsis* sucrose transporters in the split-ubiquitin system. *BMC Biochem.* **4**, 3 (2003).
18. Kronzucker, H. J., Siddiqi, M. Y. & Glass, A. Kinetics of NH_4^+ influx in spruce. *Plant Physiol.* **110**, 773–779 (1996).
19. Rawat, S. R., Silim, S. N., Kronzucker, H. J., Siddiqi, M. Y. & Glass, A. D. AtAMT1 gene expression and NH_4^+ uptake in roots of *Arabidopsis thaliana*: evidence for regulation by root glutamine levels. *Plant J.* **19**, 143–152 (1999).
20. Marini, A. M., Soussi-Boudekou, S., Vissers, S. & Andre, B. A family of ammonium transporters in *Saccharomyces cerevisiae. Mol. Cell. Biol.* **17**, 4282–4293 (1997).
21. Ludewig, U., von Wiren, N. & Frommer, W. B. Uniport of NH_4^+ by the root hair plasma membrane ammonium transporter LeAMT1;1. *J. Biol. Chem.* **277**, 13548–13555 (2002).
22. Ninnemann, O., Jauniaux, J. C. & Frommer, W. B. Identification of a high affinity NH_4^+ transporter from plants. *EMBO J.* **13**, 3464–3471 (1994).
23. Javelle, A., Severi, E., Thornton, J. & Merrick, M. Ammonium sensing in *Escherichia coli.* Role of the ammonium transporter AmtB and AmtB–GlnK complex formation. *J. Biol. Chem.* **279**, 8530–8538 (2004).
24. Marini, A. M., Boeckstaens, M., Benjelloun, F., Cherif-Zahar, B. & Andre, B. Structural involvement in substrate recognition of an essential aspartate residue conserved in Mep/Amt and Rh-type ammonium transporters. *Curr. Genet.* **49**, 364–374 (2006).
25. Changeux, J. P. & Edelstein, S. J. Allosteric mechanisms of signal transduction. *Science* **308**, 1424–1428 (2005).
26. Törnroth-Horsefield, S. *et al.* Structural mechanism of plant aquaporin gating. *Nature* **439**, 688–694 (2006).
27. Palmgren, M. G., Sommarin, M., Serrano, R. & Larsson, C. Identification of an autoinhibitory domain in the C-terminal region of the plant plasma membrane H^+-ATPase. *J. Biol. Chem.* **266**, 20470–20475 (1991).
28. Curran, A. C. *et al.* Autoinhibition of a calmodulin-dependent calcium pump involves a structure in the stalk that connects the transmembrane domain to the ATPase catalytic domain. *J. Biol. Chem.* **275**, 30301–30308 (2000).
29. Pittman, J. K. & Hirschi, K. D. Regulation of CAX1, an *Arabidopsis* Ca^{2+}/H^+ antiporter. Identification of an N-terminal autoinhibitory domain. *Plant Physiol.* **127**, 1020–1029 (2001).
30. Levy, J. *et al.* A putative Ca^{2+} and calmodulin-dependent protein kinase required for bacterial and fungal symbioses. *Science* **303**, 1361–1364 (2004).

LETTERS

The epidermis both drives and restricts plant shoot growth

Sigal Savaldi-Goldstein[1], Charles Peto[3] & Joanne Chory[1,2]

The size of an organism is genetically determined, yet how a plant or animal achieves its final size is largely unknown. The shoot of higher plants has a simple conserved body plan based on three major tissue systems: the epidermal (L1), sub-epidermal (L2) and inner ground and vascular (L3) tissues. Which tissue system drives or restricts growth has been a subject of debate for over a century[1–4]. Here, we use dwarf, brassinosteroid biosynthesis and brassinosteroid response mutants in conjunction with tissue-specific expression of these components as tools to examine the role of the epidermis in shoot growth. We show that expression of the brassinosteroid receptor or a brassinosteroid biosynthetic enzyme in the epidermis, but not in the vasculature, of null mutants is sufficient to rescue their dwarf phenotypes. Brassinosteroid signalling from the epidermis is not sufficient to establish normal vascular organization. Moreover, shoot growth is restricted when brassinosteroids are depleted from the epidermis and brassinosteroids act locally within a leaf. We conclude that the epidermis both promotes and restricts shoot growth by providing a non-autonomous signal to the ground tissues.

Because plants are sessile, the decision to grow or to stop growing is fundamental for their survival. Leaf growth involves the coordination of cell division and expansion, which is the result of developmental programs initiated by plant hormones in response to environmental cues[5,6]. One class of growth-promoting hormones are the brassinosteroids. In their absence, *Arabidopsis* plants are dwarfs, primarily owing to defects in cell expansion[7–9]. Brassinosteroid biosynthesis, catabolism and signalling mutants are known, and the genes specifying these processes have been identified[10,11]. Thus, brassinosteroid mutants and cloned genes provide a unique opportunity to examine the growth relationship between the epidermis and inner layers of an expanding leaf.

To test the contribution of the epidermis to plant growth, we drove the expression of brassinosteroid biosynthesis and perception genes under the control of the *Arabidopsis thaliana ML1* promoter in their corresponding dwarf mutant backgrounds. The *ML1* promoter is known to have highly specific expression confined to the outer L1 layer of the growing shoot[12]. We transformed the null biosynthetic *cpd* mutant (lacking detectable brassinosteroids) with the brassinosteroid biosynthetic *CPD* gene fused to yellow fluorescent protein (YFP) to generate independent *cpd;ML1-CPD–YFP* transgenic lines. As expected from the L1-specific promoter, the YFP signal was detected exclusively in cells of the outer layer (first two leaf primordia are shown, Fig. 1a). Although *CPD* is normally expressed in both the epidermis and ground tissues (Supplementary Fig. 1), *cpd;ML1-CPD–YFP* lines showed a fully rescued growth phenotype, from seedling to the adult stage (young seedlings are shown, Fig. 1d).

The ability of brassinosteroid production in the epidermis to rescue *cpd* growth defects could result from the movement of brassinosteroids to inner layers where they would be recognized by BRI1,
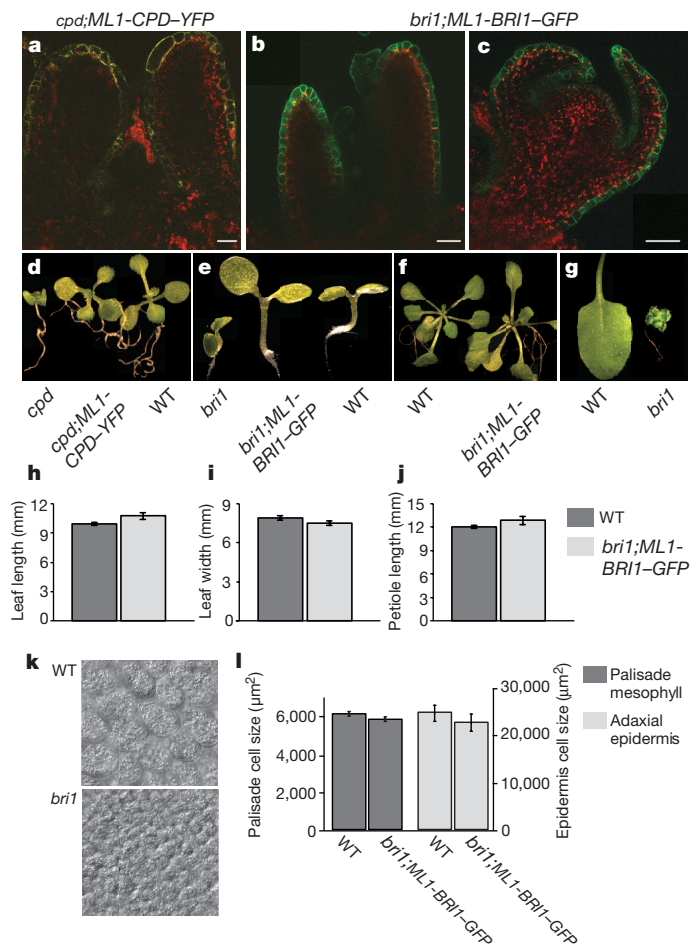
Figure 1 | Brassinosteroid biosynthesis and perception in the epidermis is sufficient to rescue dwarfism. a, **b**, **c**, CPD–YFP (**a**) and BRI1–GFP (**b**, **c**) fusion proteins accumulate specifically in the epidermis of *cpd* and *bri1* rescued plants, respectively, when expressed from the *Arabidopsis thaliana ML1* promoter. Confocal laser scanning microscopy of first two leaf primordia (**a**, **b**) and floral meristem (**c**). The images are false-coloured to indicate YFP (yellow), GFP (green) and chlorophyll auto-fluorescence (red). Scale bars, 20 μm. **d–f**, Phenotypes of representative *cpd;ML1-CPD–YFP* (**d**) and *bri1;ML1-BRI1–GFP* (**e**, **f**) transgenic lines. **g**, An adult *bri1* plant shown together with a wild-type (WT) leaf. **h–j**, Measurements indicating leaf length, leaf width and petiole length of wild type and a representative *bri1;ML1-BRI1–GFP* line. Results are presented as means ± s.e.m. (*n* > 10 plants). **k**, Light micrograph of wild-type and *bri1* mesophyll cells photographed at the same magnification. **l**, *bri1;ML1-BRI1–GFP* and wild type have similar adaxial epidermal and palisade cell area. Results are presented as means ± s.e.m. (epidermal cells, *n* = 30; palisade cells, *n* > 175).

[1]Plant Biology Laboratory, [2]Howard Hughes Medical Institute, and [3]Laboratory of Neuronal Structure and Function, The Salk Institute for Biological Studies, La Jolla, California 92037, USA.

which is expressed in both epidermal and sub-epidermal tissues in leaf cells (Supplementary Fig. 1). Thus, to limit brassinosteroid perception to the epidermis, we drove expression of the brassinosteroid receptor, BRI1, as a green fluorescent protein (GFP) fusion protein from the *ML1* promoter in a null mutant of *bri1*, *bri1-116* (first true leaf primordia and floral meristem are shown, Fig. 1b, c, respectively). In all transgenic lines analysed ($n > 20$), almost every aspect of the dwarf *bri1-116* shoot phenotype was rescued to at least wild-type stature, including hypocotyl length and leaf and petiole size, as well as the inflorescence stem height (Fig. 1e–j). The cell area in both the epidermis and the inner mesophyll layers was rescued to wild-type size (Fig. 1k, l). However, some morphological characteristics were only partially rescued (observed as a slight reduction in apical dominance and fertility, and enhanced leaf serration, data not shown).

To understand better the response of the inner layers to brassinosteroid signalling from the L1, we examined sections of fully expanded leaves and stems. Wild-type and *ML1-BRI1–GFP* leaves looked very similar (not shown). A cross section at the base of the inflorescence stem revealed fewer vascular bundles with irregular spacing in *bri1* compared with wild type, as reported previously[13]. The defect in vascular number was corrected in *bri1;ML1-BRI1–GFP* lines but these bundles exhibited defective organization and were irregularly spaced (Fig. 2a–c). Moreover, additional bundles were observed occasionally, but they were not fully developed (Fig. 2c, d). Thus, vascular organization seems to depend also on BRI1 expression in the ground layers[14]. What determines the number of vascular bundles in the stem is currently unknown. However, it has been suggested that their numbers are affected by the size of the shoot apex or by impaired auxin transport[15]. Here we confirm a role for brassinosteroids in vascular patterning, and show that BRI1's role in growth can be uncoupled from its role in vascular differentiation and organization.

The finding that a brassinosteroid signal in the epidermis was sufficient to drive growth raised the question of whether the epidermis was also necessary for the process. To address this idea, we depleted active brassinosteroids from epidermal cells by expressing the brassinosteroid catabolic enzyme, BAS1, as a YFP fusion in the epidermis of wild-type plants (Fig. 3a, b). BAS1 inactivates and reduces the functional brassinosteroid pool in the plant[16,17]. The vast majority of independent *ML1-BAS1–YFP* lines were smaller than wild type, although the degree of growth inhibition varied with the line. A representative strong dwarf line is shown together with a weak allele of *BRI1*, *bri1-5*, and wild type (Fig. 3c). This experiment suggests that optimal growth depends on local biosynthesis of brassinosteroids and that the epidermis has a role in restricting growth, in addition to promoting it. Indeed, brassinosteroids are thought to have limited mobility between and within organs, the extent of which is still unknown[18,19].

To address further the dependence of an epidermal-derived signal on the presence of brassinosteroids synthesized locally, and the contribution of the inner layer to growth, we generated lines producing brassinosteroids or expressing BRI1 in the ground tissues. Specifically, we expressed a CPD–YFP and BRI1–GFP fusion protein under the control of the *ATHB8* promoter, which has a reported expression pattern that is restricted to inner provascular cells[20]. In our hands, the *ATHB8* promoter is expressed in the vasculature, but many lines also had a diffuse expression pattern outside the vasculature tissues (Supplementary Fig. 2). Consistent with this observation, all *cpd;ATHB8-CPD–YFP* transgenic lines examined ($n > 12$) exhibited a range of semi-dwarf phenotypes. In contrast, no rescue was observed in *bri1;ATHB8-BRI1–GFP* lines (5 out of 6 lines examined) (Fig. 3d–f). Thus, the inner vasculature tissue does not seem to contribute to plant growth.

The partial rescue of *ATHB8-CPD–YFP* lines could result from movement of brassinosteroids to the outer layers. To test this possibility, we crossed one of the semi-dwarfed *cpd;ATHB8-CPD–YFP*
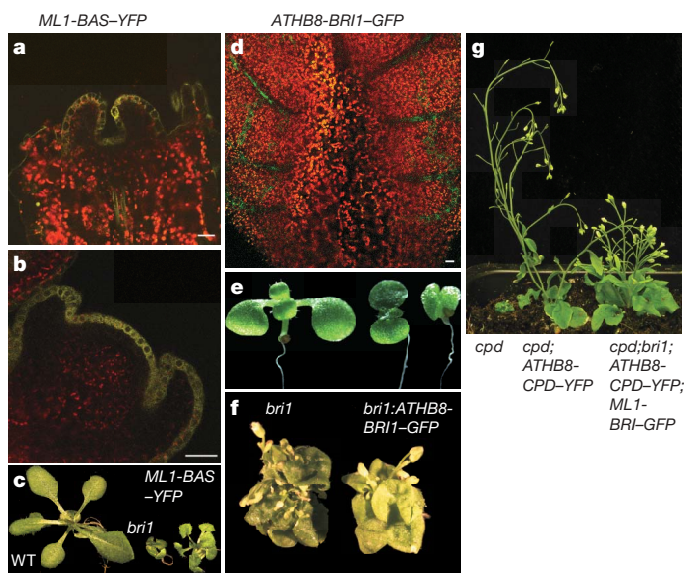

Figure 2 | Brassinosteroid signalling from the epidermis restores the number of vascular bundles but not their radial organization. Shown are cross-sections of the lower 5 mm of the inflorescence stems. **a, b,** *bri1-116* has two bundles compared with eight in wild type. **c,** Cross-section of a representative *bri1;ML1-BRI1–GFP* line. Note that bundles are not evenly distributed and sometimes clustered with each other. Some bundles are formed but they are not fully developed (marked by asterisk). **d,** Magnification of **c.** Xylem (X) and phloem (P) are indicated. Scale bars, 100 μm.

Figure 3 | Brassinosteroids act locally and make a small contribution to growth from the inner layers. **a, b,** Expression of BAS1–YFP in the L1 layer of wild-type plants restricts growth. Confocal laser scanning microscopy of the first two leaf primordia (**a**) and a floral meristem (**b**) is shown. **c,** Phenotype of wild type, *bri1-5* and wild-type plants expressing *ML1-BAS1–YFP*. **d–f,** Expression of BRI1–GFP in the vasculature can not drive growth. Confocal laser scanning microscopy of young leaf is shown (**d**). Phenotypes of young wild type, *bri1* and *bri1;ATHB8-BRI1–GFP* plants are shown (**e**); older plants are shown in **f. g,** Phenotypes of *cpd*, *cpd* expressing *ATHB8-CPD–YFP*, and the *cpd;bri* double mutant expressing *ATHB8-CPD–YFP* and *ML1-BRI1–GFP*. Scale bars, 20 μm.

lines to *bri1;ML1-BRI1–GFP*, where BRI1 is expressed from the epidermis (Fig. 3g), and selected the double mutant, *cpd;bri1*, expressing both *ATHB8-CPD–YFP* and *ML1-BRI1–GFP*. These lines typically exhibited more severe defects than the *cpd;ATHB8-CPD–YFP* lines (where endogenous BRI1 is broadly expressed). In these lines, the rosette leaves were wider (a trait associated with reduced brassinosteroid signalling), more compact and curled, and the inflorescence stem was shorter and the plants were bushier than *cpd;ATHB8-CPD–YFP* plants. With the caveat that the *ATHB8* promoter is somewhat leaky, this experiment suggests that sub-epidermal tissues may also make a small contribution to plant growth, that brassinosteroids have limited mobility and that optimal levels of active brassinosteroids are required in the epidermis for maximal growth restoration.

Our results suggest that there is a non-autonomous signal from the epidermis that triggers growth of the ground tissues. In plants, non-autonomous signalling events have been attributed to transcription factor movement through plasmodesmata[21,22]. BES1 and BZR1 transcription factors act specifically in the brassinosteroid pathway and promote the expression of growth-regulated genes[10]. A specific mutation in the coding region causes BES1 or BZR1 to accumulate to high levels and leads to BRI1-independent growth responses[23,24]. We therefore tested whether expression of a stabilized BES1 or BZR1 in the epidermis could mediate BRI1-independent growth and if so, whether this involves their intercellular movement. We generated *bri1-5* expressing *ML1-bes1-D–GFP* or *ML1-bzr1-D–GFP* (Fig. 4a).
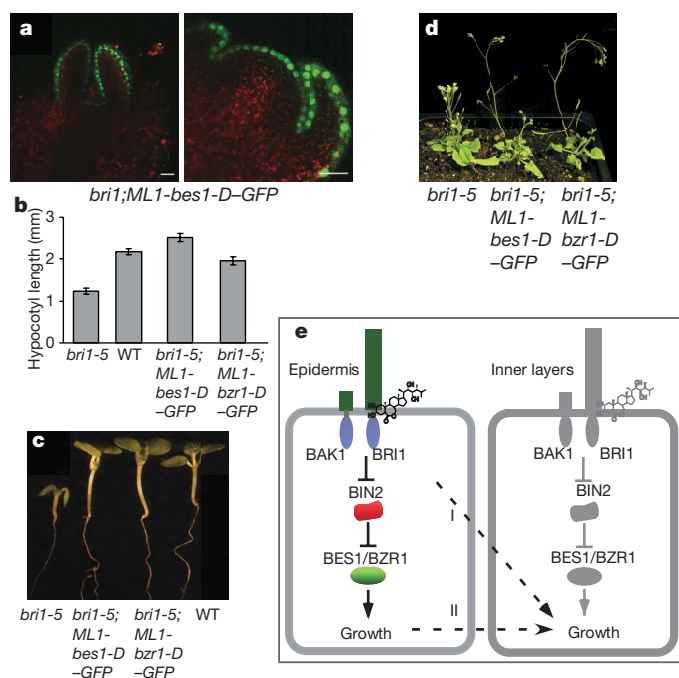
Figure 4 | The stabilized form of BES1 or BZR1 can mediate growth responses from the epidermis in *bri1*. a, Expression of *bes1-D–GFP* in the epidermis of the first two leaf primordia (left panel) and floral meristem (right panel) in *bri1-5*. Scale bars, 20 μm. **b,** Hypocotyl length of 5-day-old *bri1-5*, *bri1-5;ML1-bes1-D–GFP*, *bri1-5;ML1-bzr1-D–GFP* and wild-type Wassilewskija. Results are presented as means ± s.e.m. ($n > 12$). **c,** Phenotypes of *bri1-5*, *bri1-5;ML1-bes1-D–GFP*, *ML1-bzr1-D–GFP* seedlings and their corresponding wild-type background. **d,** *ML1-bes1-D–GFP* and *ML1-bzr1-D–GFP* at adult stage. **e,** A model illustrating the contribution of the epidermis to growth. A simplified summary of brassinosteroid signalling is shown. BRI1 binds brassinosteroids at the plasma membrane and interacts with its proposed co-receptor BAK1, which relieves BIN2's negative regulation of BES1/BZR1. In the presence of brassinosteroids, BES1/BZR1 can bind to DNA, thus regulating the expression of hundreds of genes, many of which are involved in growth. The expanding epidermis can signal non-autonomously to the inner layers. The signal could be linked to the brassinosteroid signalling pathway (I) or could result from the growth process itself (for example, mechanical stimulus; II).

The absence of detectable GFP signal of the highly expressed bes1-D–GFP protein in the inner layers emphasizes the specificity of the promoter, as well as the inability of bes1-D–GFP protein to move to the next layer (similar results were obtained for bzr1-D–GFP; data not shown and supplementary Fig. 2). As opposed to the short hypocotyls of *bri1-5*, *bri1;ML1-bes1-D–GFP* and *bri1;ML1-bzr1-D–GFP* lines had larger cotyledons and long hypocotyls (Fig. 4b, c). At the adult stage, enhanced growth was also observed in centric organs such as pedicels, internodes and the overall height of the inflorescence stem, albeit not to the same extent as compared with wild type (Fig. 4d). The leaf phenotypes of *bri1;ML1-bes1-D–GFP* or *bri1;ML1-bzr1-D–GFP* lines were different in size and shape from both the wild-type ecotype and *bri1-5*, with a range of leaf shapes from wide leaves to longer and slightly curled ones (aberrant leaf form is typical to *bes1-D* and *bzr1-D* mutants[23,24]). Thus, some aspects of the non-autonomous growth signal are generated downstream of BES1 and BZR1. The incomplete rescue by BES1/BZR1 expression may reflect the existence of additional epidermal components required for the pathway.

Taken together, our data suggest that brassinosteroid perception in the epidermis has a major role in both driving and restricting shoot growth. Although the dramatic rescue of dwarfism when BRI1 is expressed from the L1 occurs by an as yet unknown mechanism, the non-autonomous signal seems to be unidirectional—coming from the outer to inner layers. Yet, it would be hard to completely exclude a role for internal tissues in promoting plant growth. The growth of cells in the inner layers may be a response to a moving brassinosteroid-generated signal that is neither BES1 nor BZR1, the two known endpoints of the signalling pathway. Alternatively, the inner layers could passively or actively follow the expansion of the epidermis by means of mechanical stimuli (Fig. 4e). Recently, mechanical wall relaxation applied locally to the shoot apex was shown to induce growth of a leaf primordium[25]. The development of mechanical markers and brassinosteroid sensors would help to distinguish these possibilities.

METHODS

Plasmid constructs for transgenic lines. All final constructs were in pMLBART binary vector[26]. The *ML1* promoter (fragment 2, as described[12]) was cloned as a *Xho*I/*Kpn*1 fragment to the polylinker of pBJ36[26]. A fragment of 1.7 kilobases of the *ATHB8* genomic sequence was cloned by *Pst*I/*Kpn*I into pBJ36[20]. *BRI1–GFP*, *bes1-D–GFP* and *bzr1-D–GFP* constructs were cloned as *Kpn*1 fragments, and *BAS1–YFP* and *CPD–YFP* were cloned as *Sma*I/*Bam*HI fragments to BJ36/*ML1* or BJ36/*ATHB8*.

Plant stocks and growth conditions. *bri1-116* and *bri1-5* are in the Columbia (Col-0) and Wassilewskija (WS) backgrounds, respectively. *cpd* was isolated as described in ref. 7. Plants were grown either on soil under greenhouse conditions or in half-strength Linsmaier–Skoog medium (Caisson Laboratories) on a 16 h light/8 h dark cycle.

Plant cross sections, measurements and microscopy. Stems were fixed in 1.25% glutaraldehyde in 0.05 M sodium cacodylate buffer (Electron Microscopy Sciences, Hatfield, Pennsylvania), dehydrated in ethanol, embedded in Technovit 7100, sectioned at 3 μm, and stained with toluidine blue. For cell measurements, fully expanded leaf number 3 or 4 from 5-week-old plants were cleared in 100% methanol followed by 90% lactic acid. Cell area was measured using the ImageJ software. Hypocotyl and leaf lengths were measured with NIH Image 1.62. Olympus BX60 or Leica DM5000B compound microscopes were used for light microscopy. Fluorescence signals were detected using a Leica SP 2 confocal microscope.

1. Peters, W. S. & Tomos, D. The epidermis still in control? *Bot. Acta* 109, 264–267 (1996).
2. Kutschera, U. Stem elongation and cell wall proteins in flowering plants. *Plant Biol.* 3, 466–480 (2001).
3. Green, P. B. Connecting gene and hormone action to form, pattern and organogenesis: biophysical transductions. *J. Exp. Bot.* 45, 1775–1788 (1994).
4. Dale, J. E. The control of leaf expansion. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 39, 267–295 (1988).
5. Tsukaya, H. Mechanism of leaf-shape determination. *Annu. Rev. Plant Biol.* 57, 477–496 (2006).

6. Gendreau, E. *et al.* Cellular basis of hypocotyl growth in *Arabidopsis thaliana. Plant Physiol.* **114**, 295–305 (1997).

7. Szekeres, M. *et al.* Brassinosteroids rescue the deficiency of CYP90, a cytochrome P450, controlling cell elongation and de-etiolation in *Arabidopsis. Cell* **85**, 171–182 (1996).

8. Nakamura, A. *et al.* The role of *OsBRI1* and its homologous genes, *OsBRL1* and *OsBRL3*, in rice. *Plant Physiol.* **140**, 580–590 (2006).

9. Nakaya, M., Tsukaya, H., Murakami, N. & Kato, M. Brassinosteroids control the proliferation of leaf cells of *Arabidopsis thaliana. Plant Cell Physiol.* **43**, 239–244 (2002).

10. Vert, G., Nemhauser, J. L., Geldner, N., Hong, F. & Chory, J. Molecular mechanisms of steroid hormone signaling in plants. *Annu. Rev. Cell Dev. Biol.* **21**, 177–201 (2005).

11. Fujioka, S. & Yokota, T. Biosynthesis and metabolism of brassinosteroids. *Annu. Rev. Plant Biol.* **54**, 137–164 (2003).

12. Sessions, A., Weigel, D. & Yanofsky, M. F. The *Arabidopsis thaliana MERISTEM LAYER 1* promoter specifies epidermal expression in meristems and young primordia. *Plant J.* **20**, 259–263 (1999).

13. Choe, S. *et al.* The *Arabidopsis dwf7/ste1* mutant is defective in the delta7 sterol C-5 desaturation step leading to brassinosteroid biosynthesis. *Plant Cell* **11**, 207–221 (1999).

14. Cano-Delgado, A. *et al.* BRL1 and BRL3 are novel brassinosteroid receptors that function in vascular differentiation in *Arabidopsis. Development* **131**, 5341–5351 (2004).

15. Turner, S. & Sieburth, L. E. Vascular Patterning. In *The Arabidopsis Book* (eds Somerville, C. R. & Meyerowitz, E. M.) (Am. Soc. Plant Biol., Rockville, Maryland, 2003).

16. Turk, E. M. *et al.* CYP72B1 inactivates brassinosteroid hormones: an intersection between photomorphogenesis and plant steroid signal transduction. *Plant Physiol.* **133**, 1643–1653 (2003).

17. Neff, M. M. *et al. BAS1*: A gene regulating brassinosteroid levels and light responsiveness in *Arabidopsis. Proc. Natl Acad. Sci. USA* **96**, 15316–15323 (1999).

18. Symons, G. M. & Reid, J. B. Brassinosteroids do not undergo long-distance transport in pea. Implications for the regulation of endogenous brassinosteroid levels. *Plant Physiol.* **135**, 2196–2206 (2004).

19. Bishop, G. J., Harrison, K. & Jones, J. D. The tomato *Dwarf* gene isolated by heterologous transposon tagging encodes the first member of a new cytochrome P450 family. *Plant Cell* **8**, 959–969 (1996).

20. Baima, S. *et al.* The expression of the *Athb-8* homeobox gene is restricted to provascular cells in *Arabidopsis thaliana. Development* **121**, 4171–4182 (1995).

21. Kurata, T., Okada, K. & Wada, T. Intercellular movement of transcription factors. *Curr. Opin. Plant Biol.* **8**, 600–605 (2005).

22. Gallagher, K. L. & Benfey, P. N. Not just another hole in the wall: understanding intercellular protein trafficking. *Genes Dev.* **19**, 189–195 (2005).

23. Yin, Y. *et al.* BES1 accumulates in the nucleus in response to brassinosteroids to regulate gene expression and promote stem elongation. *Cell* **109**, 181–191 (2002).

24. Wang, Z. Y. *et al.* Nuclear-localized BZR1 mediates brassinosteroid-induced growth and feedback suppression of brassinosteroid biosynthesis. *Dev. Cell* **2**, 505–513 (2002).

25. Pien, S., Wyrzykowska, J., McQueen-Mason, S., Smart, C. & Fleming, A. Local expression of expansin induces the entire process of leaf development and modifies leaf shape. *Proc. Natl Acad. Sci. USA* **98**, 11812–11817 (2001).

26. Gleave, A. P. A versatile binary vector system with a T-DNA organisational structure conducive to efficient integration of cloned DNA into the plant genome. *Plant Mol. Biol.* **20**, 1203–1207 (1992).

LETTERS

C5L2 is critical for the biological activities of the anaphylatoxins C5a and C3a

Nien-Jung Chen[1], Christine Mirtsos[1], Daniel Suh[1], Yong-Chen Lu[1], Wen-Jye Lin[1], Colin McKerlie[2], Taeweon Lee[3], Helene Baribault[3], Hui Tian[3] & Wen-Chen Yeh[1,3]

Complement-derived anaphylatoxins regulate immune and inflammatory responses through G-protein-coupled receptor (GPCR)-mediated signalling[1–4]. C5L2 (also known as GPR77) is a relatively new GPCR thought to be a non-signalling receptor binding to C5a, on the basis of sequence information and experimental evidence[5–7]. Here we show, using gene targeting, that C5L2 is required to facilitate C5a signalling in neutrophils, macrophages and fibroblasts *in vitro*. Deficiency of C5L2 results in reduced inflammatory cell infiltration, suggesting that C5L2 is critical for optimal C5a-mediated cell infiltration in certain *in vivo* settings. C5L2 is also involved in optimizing C3a-induced signals. Furthermore, like mice incapable of C3a/complement 3a receptor (C3aR) signalling[4,8,9], C5L2-deficient mice are hypersensitive to lipopolysaccharide (LPS)-induced septic shock, show reduced ovalbumin (OVA)-induced airway hyperresponsiveness and inflammation, and are mildly delayed in haematopoietic cell regeneration after γ-irradiation. Our data indicate that C5L2 can function as a positive modulator for both C5a- and C3a-anaphylatoxin-induced responses.

Proteolysis of C3 and C5 are central events of complement activation. The smaller fragments, C3a and C5a, are known anaphylatoxins and induce or augment several biological responses. C3aR[10], C5aR[11] and C5L2 share significant homology and belong to the GPCR family. C5a and C5a des-Arg bind C5L2 with high affinity[6,12,13], but whether or not C3a and C4a are ligands for C5L2 remains debatable[12–14]. We generated C5L2-deficient mice (Supplementary Information section 1), and reverse-transcriptase–PCR (RT–PCR) analysis showed that although C5aR expression was comparable in $c5l2^{-/-}$ and wild-type livers, C5L2 messenger RNA was present only in wild-type liver (Fig. 1a). Additionally C5L2 is detectable on wild-type but not on $c5l2^{-/-}$ neutrophils (Fig. 1b) and macrophages (Supplementary Information section 1). C5aR and C3aR protein levels were comparable in $c5l2^{-/-}$ and wild-type neutrophils, using antibodies obtained commercially as described (Supplementary Information section 1). $c5l2^{-/-}$ mice appeared healthy and displayed no obvious developmental abnormalities.

The binding of mouse C5L2 to mouse C5a was first confirmed using COS cells overexpressing mouse C5L2 (Fig. 1c, d). To investigate the role of C5L2 in C5a-mediated responses, neutrophils were stimulated and the surface Mac-1 (also known as Itgam) induction was monitored; Mac-1 is normally enhanced when neutrophils are activated. $c5l2^{-/-}$ neutrophils showed impaired Mac-1 induction compared with wild-type cells in response to C5a alone (Fig. 1e, left). This defect was magnified when neutrophils were stimulated with C5a plus LPS and was also evident in a BALB/c background (Fig. 1e, right). We also examined TNF-α and IL-6 production by neutrophils after stimulation. LPS alone induced comparable levels of TNF-α

(Fig. 1f, left) and IL-6 (Fig. 1f, right) in wild-type and $c5l2^{-/-}$ neutrophils. As previously reported, C5a treatment reduced TNF-α production[15], but enhanced IL-6 production[16] in LPS-stimulated wild-type neutrophils. In contrast, C5a was unable to affect LPS-induced TNF-α and IL-6 production in $c5l2^{-/-}$ cells. Furthermore, C5a des-Arg induced strong Mac-1 induction in wild-type neutrophils, but not in $c5l2^{-/-}$ cells. However, neither C3a nor C3a des-Arg induced Mac-1 expression in wild-type or $c5l2^{-/-}$ neutrophils (Supplementary Information section 2).

MAPK and PKB/Akt activation are key downstream events induced by C5a[16,17]. In wild-type neutrophils, C5a treatment alone strongly induced ERK1/2 activation, and weakly enhanced phospho-p38 expression (Fig. 2a). A low dose (20 ng ml^{-1}) of LPS enhanced C5a-induced activations of various MAPKs. In contrast, $c5l2^{-/-}$ neutrophils showed only weak induction of activated ERK1/2, JNK or p38 in response to C5a, either alone or with LPS.

C5a is also known to have a critical role in regulating macrophage functions[18]. We found that C5a-induced activation of ERK1/2 and AKT was severely impaired in C57BL/6 $c5l2^{-/-}$ macrophages (Fig. 2b) and BALB/c $c5l2^{-/-}$ macrophages (Fig. 2c). In addition, LPS strongly upregulated expression of the co-stimulation molecules CD40 and CD86 in wild-type macrophages, whereas C5a suppressed this induction. In contrast, $c5l2^{-/-}$ macrophages showed slightly higher basal CD40 expression, but a substantially lower level of LPS-induced CD40 and CD86 expression. Furthermore, C5a-mediated suppression of LPS-induced CD40 and CD86 expression was impaired in $c5l2^{-/-}$ macrophages (Fig. 2d; Supplementary Information section 2). Unlike in neutrophils, C5a suppressed LPS-induced TNF-α and IL-6 production in wild-type macrophages: a process that is impaired in $c5l2^{-/-}$ cells (Fig. 2e). Taken together, C5L2 is an important downstream mediator of C5a functions. Furthermore, specific aspects of LPS signalling seem to depend on the presence of C5L2.

To determine whether C5L2 is involved in C3a-initiated signalling, we stimulated neutrophils with C3a, and examined downstream signals. Phosphorylated ERK1/2 and AKT were readily detected in wild-type neutrophils, following stimulation, but these signals were substantially reduced in C3a-stimulated $c5l2^{-/-}$ neutrophils (Fig. 3a). Similar reductions were detected in bone marrow derived $c5l2^{-/-}$ macrophages (Supplementary Information section 3). Actin polymerization is an early consequence of neutrophil activation induced by chemoattractants. We detected C3a-induced F-actin formation in wild-type and $c5l2^{-/-}$ neutrophils (Fig. 3b; Supplementary Information section 3). Almost 50% of wild-type cells were positive for F-actin staining after stimulation, compared with only 11% of C3a-stimulated $c5l2^{-/-}$ neutrophils. Thus, C5L2 is also required for optimal C3a signalling.

[1]The Campbell Family Institute for Breast Cancer Research, Ontario Cancer Institute, University Health Network and Department of Medical Biophysics, University of Toronto, Toronto, Ontario M5G 2C1, Canada. [2]Hospital for Sick Children, Toronto, Ontario M5G 1X8, Canada. [3]Amgen San Francisco, 1120 Veterans Boulevard, South San Francisco, California 94080, USA.

C5a/C5aR-mediated signalling is important for regulation of inflammatory responses *in vivo*[3,19]. We investigated the effect of C5L2 deficiency on thioglycollate (TGC)-induced peritonitis. In comparison to TGC-treated wild-type mice, TGC-treated $c5l2^{-/-}$ mice showed a mild reduction (35%) in macrophage infiltration into the peritoneal cavity at 72 h (Fig. 4a). Consistent with a recent report[7], $c5l2^{-/-}$ mice were capable of responding to intra-peritoneal injection of recombinant C5a and no significant difference was observed between wild-type and mutant mice in terms of short-term neutrophil recruitment (data not shown). This suggests that absence of C5L2 does not impact on *in vivo* C5a signals to the same degree as C5aR deficiency, and that certain short-term cell recruitment effects require only C5aR.

To characterize further longer-term defects of inflammatory cell migration in $c5l2^{-/-}$ mice, we generated dorsal air pouches on mice six days before injection. C5a (2 μg) combined with 100 μl TGC in 1.5 ml CMC solution, were injected into the air pouches. Infiltrated

cells were recovered and analysed for total cell numbers and numbers of neutrophils and macrophages. CMC or TGC injection alone induced similar levels of cell infiltration in wild-type and $c5l2^{-/-}$ mice. Injection of C5a alone into wild-type air pouches induced a modest level of infiltration for total cells, neutrophils and macrophages. A mild deficit of 37% in these activities was apparent in $c5l2^{-/-}$ air pouches compared with wild type (Supplementary Information section 4). Co-injection of TGC and C5a induced inflammatory cell infiltration that was more robust in wild-type air pouches versus those of $c5l2^{-/-}$ mice. Fewer neutrophils (2.3-fold) and fewer macrophages (3.5-fold) were present in the inflammatory infiltrates of TGC plus C5a-treated $c5l2^{-/-}$ mice compared with wild-type controls 24 h after stimulation (Supplementary Information section 4). Importantly, activation of the infiltrated neutrophils in the air pouches of $c5l2^{-/-}$ mice was impaired compared with wild-type mice (Fig. 4b).

C5a- and C3a-mediated immune responses have been implicated in asthma pathogenesis. We therefore examined how C5L2 deficiency affects the development of asthma-like airway hyper-responsiveness and inflammation using the commonly employed mouse model induced by multiple OVA challenges in a BALB/c background. Airway hyper-responsiveness to inhaled methacholine in OVA-sensitized wild-type mice is markedly enhanced compared with the normal saline treated controls. Conversely, methacholine-induced airway hyper-responsiveness in OVA-sensitized $c5l2^{-/-}$ mice was markedly reduced compared with the sensitized wild-type littermates (Fig. 4c). Histological analyses also demonstrated that allergen-induced inflammatory cell infiltration was markedly reduced in $c5l2^{-/-}$ mice versus wild-type littermates (Fig. 4d; Supplementary
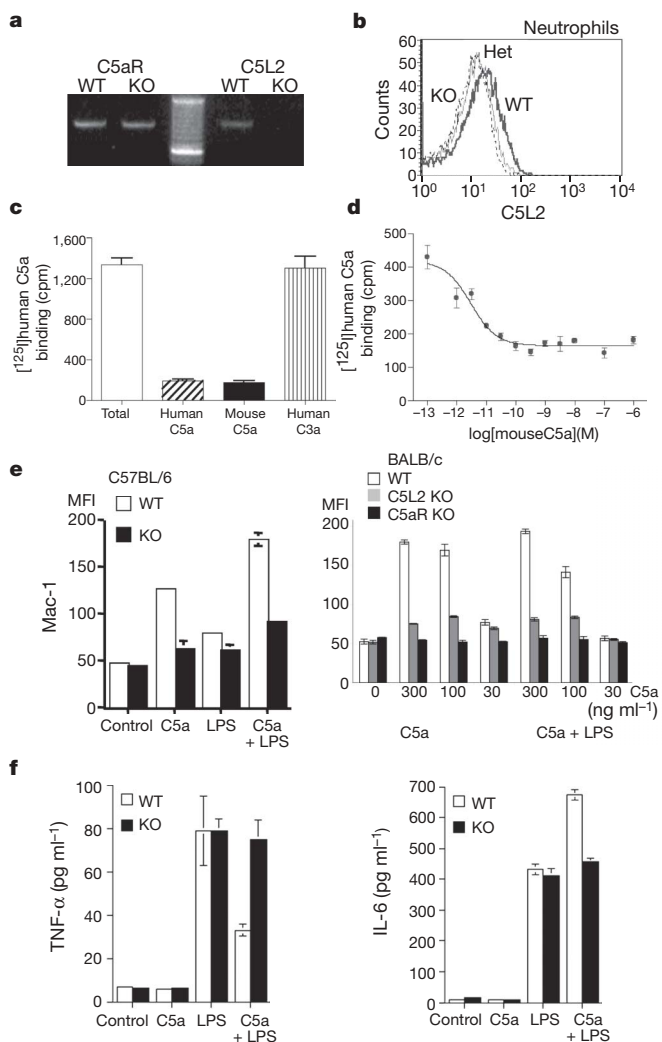
Figure 1 | C5L2 expression and C5a activation in $c5l2^{-/-}$ neutrophils.
a, RT–PCR analysis of C5L2 and C5aR mRNA in liver tissues. **b**, C5L2 expression measured using flow cytometry and an anti-C5L2 antibody in wild-type (WT), heterozygote (Het) and $c5l2^{-/-}$ (KO) neutrophils. **c**, Inhibition of [^{125}I]humanC5a binding by human and mouse C5a (0.1 μM) to C5L2-expressing COS cell membrane (50 μg) (mean ± s.d., $n = 4$). **d**, Binding affinity of mouse C5a to mouse C5L2 (median inhibitory concentration = 3.9 ± 0.7 pM; mean ± s.d., $n = 4$). **e**, Mac-1 expression on stimulated neutrophils. Left panel, C57BL/6 background; right, BALB/c background. MFI, mean fluorescence intensity. **f**, TNF-α and IL-6 production by neutrophils 4 h after stimulation. For **e** and **f**, data are expressed as the mean ± s.d. of triplicate determinations.
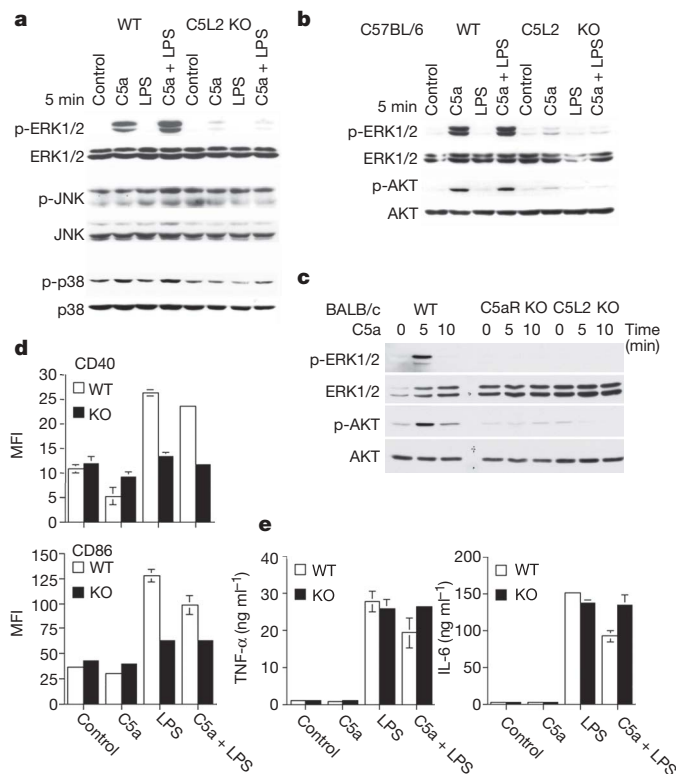
Figure 2 | Reduced C5a-induced downstream signalling in $c5l2^{-/-}$ cells.
a, Decreased MAP kinase activation in neutrophils after stimulation.
b, Reduced ERK1/2 and AKT activation in C57BL/6 macrophages. **c**, Reduced ERK1/2 and AKT activation in BALB/c $c5l2^{-/-}$ and $c5ar^{-/-}$ macrophages. **d**, Impaired induction of co-stimulatory molecules in $c5l2^{-/-}$ macrophages after stimulation. **e**, Cytokine production by wild-type and $c5l2^{-/-}$ macrophages. Results shown are from one experiment representative of two independent trials. For **d** and **e**, data are the mean ± s.d. of triplicate samples.

Information section 5). This indicates that C5L2 is involved in the pathogenesis of asthma-like airway hyper-responsiveness and inflammation, potentially owing to the inputs of both C5a and C3a signals.

We further investigated the response of C5L2-deficient mice to LPS challenge. Mice were injected intraperitoneally with 500 µg LPS and mortality was monitored. Interestingly, $c5l2^{-/-}$ mice showed increased susceptibility to lethal effects of LPS injection compared with control littermates (Fig. 4e). Furthermore, LPS-injected mutant mice showed higher IL-1β serum levels than wild-type littermates, indicating that the increased susceptibility was associated with elevation of some inflammatory cytokines (Fig. 4f). Similar results were previously reported for mice lacking C3aR-mediated signalling[8]. Thus, like C3aR, C5L2 is involved in a regulatory mechanism that normally protects against LPS-induced shock responses.

C3a also initiates signalling to promote haematopoietic stem cell engraftment[9]. We monitored the regeneration of haematopoietic cells in $c5l2^{-/-}$ mice after sub-lethal irradiation at 3 Gy, and found that peripheral blood leukocyte recovery was mildly perturbed in these mutant mice compared with wild-type controls (Fig. 4g), further supporting the idea that C5L2 may be involved in optimizing C3a-dependent signalling in vivo.

Recently, several studies demonstrated that C5L2 has functional roles in regulating inflammation[7,20] and sepsis[21]. C5L2 has also been reported to mediate ASP-induced signalling and β-arrestin redistribution[14], suggesting that C5L2 is capable of transducing signals. C5L2 may contribute to signalling induced by C5a or C3a by heterodimerizing with C5aR or C3aR. Intriguingly, we found that C5L2 may be critical for upregulating cell surface expression of C5aR particularly in neutrophils (Supplementary Information sections 6–8). This could be one of the ways in which C5L2 contributes to optimizing C5a signals. GPCR heterodimerization, for example, between GABA_B receptor subunits, could potentially promote receptor expression on the cell surface[22]. However, while the possibilities of heterodimerization still exist, we are unable yet to detect interactions between C5L2 and the classical anaphylatoxin receptors.

Another potential mechanism is that ligation of C5L2 may trigger signals and contribute to C5a (and C3a) signalling outputs in parallel to the classical receptors. Ectopic expression of C5L2 in rat glioma cells can decrease inflammatory responses even without ligand stimulation[20]. To investigate further the role of C5L2 in C5a-mediated signalling, we ectopically expressed C5L2 in $c5l2^{-/-}$ and $c5ar^{-/-}$ lung fibroblasts. Interestingly, C5a could induce a weak ERK

activation signal in $c5ar^{-/-}$ cells (Fig. 5a). This suggested that, at least in lung fibroblasts, a C5aR-independent signal exists on C5a stimulation. Overexpression of C5L2 in $c5l2^{-/-}$ cells, but not in $c5ar^{-/-}$ cells, strongly enhanced ERK activation induced by C5a stimulation (Fig. 5a). In contrast, $c5l2^{-/-}$ cells reconstituted with a truncated C5L2 without the first amino-terminal extracellular domain (noECD) or with a truncated last carboxy-terminal intracellular domain (noICD), did not show enhanced ERK signals in response to C5a (Fig. 5b). These results indicate that reconstitution of C5L2 can mediate optimal C5a signal in the $c5l2^{-/-}$ cells, and that the presence of both C5aR and C5L2 are required for optimal signals to be induced by C5a.

We generated stable RAW cell lines to assess more effects of C5L2 overexpression. RAW cells contain endogenous C5L2 and C5aR, therefore the elevated ERK activation induced by C5L2 overexpression versus that in the control was not as dramatic as in lung fibroblasts (Fig. 5c). Interestingly, overexpression of C5L2 reduced
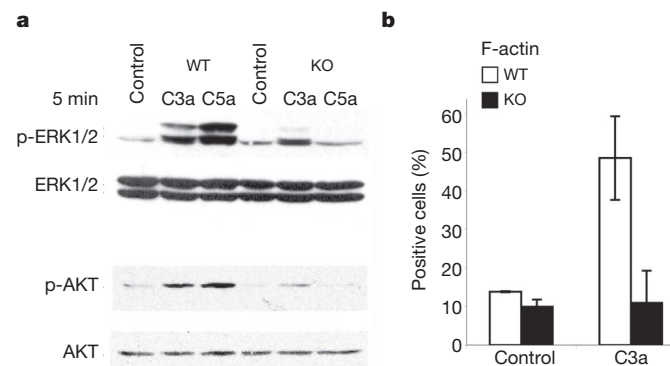

Figure 4 | Defective in vivo inflammatory responses and haematopoietic cell regeneration in $c5l2^{-/-}$ mice. a, Reduced numbers of inflammatory cells in the peritoneal cavity (mean ± s.d., n = 5). **b**, Impaired neutrophil activation in air pouches. Infiltrated cells in the air pouches were examined 24 h after stimulation (mean ± s.d., n = 3). **c**, Reduced OVA-induced asthma-like airway hyper-responsiveness and inflammation in $c5l2^{-/-}$ mice. Penh, enhanced pause—a measure for airway hyper-responsiveness; NS, normal saline. Asterisk, P-value as shown, n = 6, by two-tailed, paired, t-test. **d**, Histological analysis of OVA-induced lung tissue. Haematoxylin and eosin stained, ×100 magnification. **e**, Increased LPS-induced septic lethality (n = 8, three independent trials). **f**, Serum IL-1β levels in the mice in **d**. **g**, Delayed haematopoietic cell regeneration in $c5l2^{-/-}$ mice (n = 3, two independent trials). Error bars in **f** and **g** are s.d. PBL, peripheral blood leukocyte.

Figure 3 | Reduced C3a-induced signalling and F-actin formation in $c5l2^{-/-}$ neutrophils. a, Decreased C3a-induced ERK1/2 and AKT activation. Neutrophils were stimulated for 5 min with C3a (1 µg ml⁻¹) or C5a (100 ng ml⁻¹) and phosphorylation of ERK1/2 and AKT was determined. **b**, Decreased F-actin formation. Wild-type and $c5l2^{-/-}$ neutrophils were treated with C3a (12.5 µg ml⁻¹) for 1 min to induce F-actin formation. Data shown are the mean percentage of positive cells ± s.d. of three samples. The results shown are from one experiment representative of two independent trials.
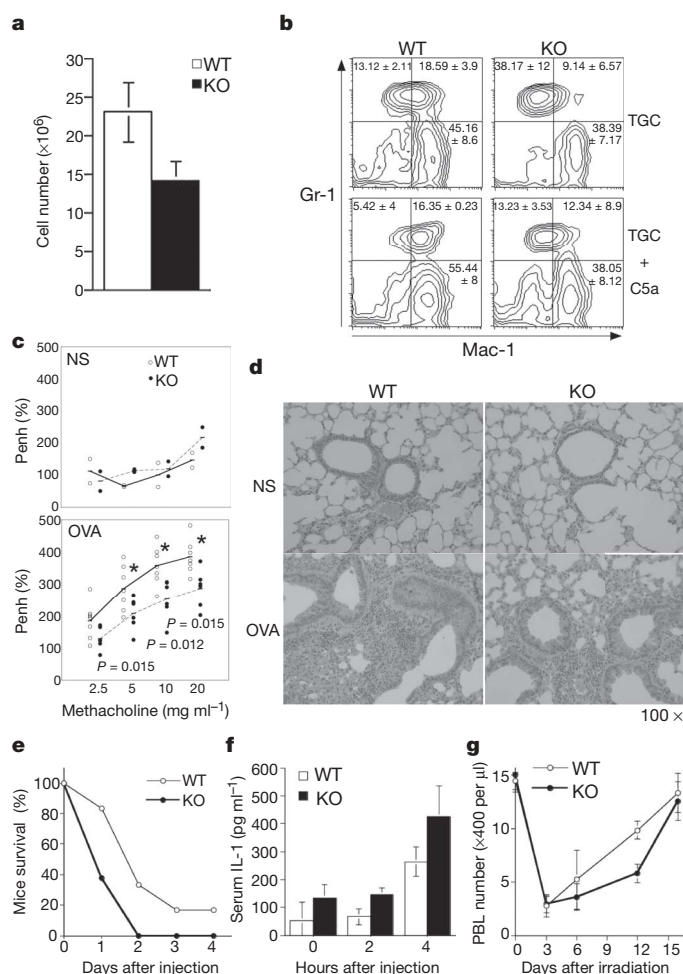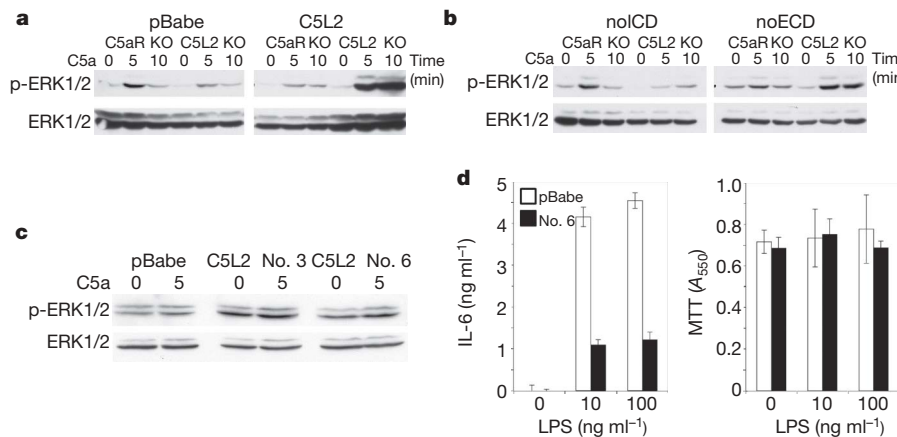
Figure 5 | Overexpression of C5L2 in lung fibroblasts and RAW cells.
a, Increased ERK1/2 activation in lung fibroblasts expressing full-length C5L2 compared with vector-control lung fibroblasts (pBabe), both with 100 ng ml^{-1} C5a stimulation. **b**, C5a-induced ERK activation in lung fibroblasts transfected with truncated C5L2 without a C-terminal intracellular domain (noICD) or without an N-terminal extracellular domain (noECD). **c**, C5a-induced ERK1/2 activation in C5L2-expressing RAW transfectants (C5L2, two independent clones) compared with vector-control RAW transfectant (pBabe). **d**, Reduced LPS-induced IL-6 production (mean ± s.d., $n = 3$) in C5L2-transfected RAW clones (left). Comparable cell numbers after 24 h. LPS stimulation in RAW cell transfectants shown by MTT (3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide) assay (mean ± s.d., $n = 3$) (right). This colorimetric assay measures cell viability and activity of a mitochondrial enzyme.

LPS-induced IL-6 production (Fig. 5d). These results are consistent with the possibility that overexpressed C5L2 can generate signals to activate ERK as well as to downregulate LPS/TLR4-induced responses.

In this report, we demonstrate that C5L2 can act as a positive regulator that is critical for C5a and C3a signalling. Our results indicate that C5L2 can work cooperatively with signals mediated by C5aR and C3aR. The involvement of C5L2 in controlling both pro-inflammatory and anti-inflammatory responses (the latter consistent with a recent report[7]) and, in particular, certain aspects of LPS signalling makes C5L2 a potential target for treating diseases such as asthma, allergy and sepsis.

METHODS

Reagents. Lipopolysaccharide (LPS from *E. coli* O55B5), Histopaque-1083, fetal bovine serum and purified human C5a was purchased from Sigma. Recombinant mouse C5a was purchased from R&D systems. Purified human C3a, C3a des-Arg and C5a des-Arg were purchased from Calbiochem. Thioglycollate was purchased from Gibco Invitrogen. Polyclonal antibodies against mouse C5aR (P14 and H100) and C3aR (E18 and D20) were purchased from Santa Cruz. Monoclonal antibody against mouse C5aR (clone 20/70) was purchased from Cedarlane. Polyclonal antibody against mouse C5L2 was purchased from HyCult biotechnology. Fluorochrome-conjugated antibodies used in cell typing in flow cytometry were purchased from BD Pharmingen and were used at optimal titrated concentration.

Generation and screening of *c5l2*$^{-/-}$ mice. The *c5l2*$^{-/-}$ mice were generated by Lexicon Genetics (The Woodlands, Texas). Details can be found in Supplementary Information.

Preparation of neutrophils, macrophages and lung fibroblasts. Neutrophils were purified from bone marrow as previously described[23]. Macrophages were purified from mice 3 days after they were intraperitoneally injected with 1 ml of TGC broth (2.4% in H$_2$O). Lung fibroblasts were cultured from adult mice lung tissues. Details can be found in Supplementary Information.

Measurement of cytokine levels. Levels of IL-1β, TNF-α and IL-6 in sera or culture supernatants were measured with ELISA kits from BD bioscience, following the manufacturer's instructions.

Assessment of *in vitro* cellular activation. Surface activation markers (CD40, CD86, Gr-1 and Mac-1) were detected by flow cytometry. Activation of MAP kinases and AKT was examined by western blotting. C3a-induced F-actin formation was detected by phalloidin staining and confocal microscopy. Details can be found in Supplementary Information.

Assessment of *in vivo* immune activation in mice. Details of acute peritonitis, air pouches[24], the OVA-induced asthma mouse model and haematopoietic cell regeneration experiments are described in Supplementary Information.

1. Shin, H. S., Snyderman, R., Friedman, E., Mellors, A. & Mayer, M. M. Chemotactic and anaphylatoxic fragment cleaved from the fifth component of guinea pig complement. *Science* **162**, 361–363 (1968).
2. Bokisch, V. A., Muller-Eberhard, H. J. & Cochrane, C. G. Isolation of a fragment (C3a) of the third component of human complement containing anaphylatoxin and chemotactic activity and description of an anaphylatoxin inactivator of human serum. *J. Exp. Med.* **129**, 1109–1130 (1969).
3. Höpken, U. E., Lu, B., Gerard, N. P. & Gerard, C. The C5a chemoattractant receptor mediates mucosal defence to infection. *Nature* **383**, 86–89 (1996).
4. Humbles, A. A. *et al.* A role for the C3a anaphylatoxin receptor in the effector phase of asthma. *Nature* **406**, 998–1001 (2000).
5. Ohno, M. *et al.* A putative chemoattractant receptor, C5L2, is expressed in granulocyte and immature dendritic cells, but not in mature dendritic cells. *Mol. Immunol.* **37**, 407–412 (2000).
6. Cain, S. A. & Monk, P. N. The orphan receptor C5L2 has high affinity binding sites for complement fragments C5a and C5a des-Arg^{74}. *J. Biol. Chem.* **277**, 7165–7169 (2002).
7. Gerard, N. P. *et al.* An anti-inflammatory function for the complement anaphylatoxin C5a-binding protein, C5L2. *J. Biol. Chem.* **280**, 39677–39680 (2005).
8. Kildsgaard, J. *et al.* Cutting edge: targeted disruption of the C3a receptor gene demonstrates a novel protective anti-inflammatory role for C3a in endotoxin-shock. *J. Immunol.* **165**, 5406–5409 (2000).
9. Ratajczak, M. Z. *et al.* Transplantation studies in C3-deficient animals reveal a novel role of the third complement component (C3) in engraftment of bone marrow cells. *Leukemia* **18**, 1482–1490 (2004).
10. Hsu, M. H. *et al.* Cloning and functional characterization of the mouse C3a anaphylatoxin receptor gene. *Immunogenetics* **47**, 64–72 (1997).
11. Perret, J. J., Raspe, E., Vassart, G. & Parmentier, M. Cloning and functional expression of the canine anaphylatoxin C5a receptor. Evidence for high interspecies variability. *Biochem. J.* **288**, 911–917 (1992).
12. Kalant, D. *et al.* The chemoattractant receptor-like protein C5L2 binds the C3a des-Arg^{77}/acylation-stimulating protein. *J. Biol. Chem.* **278**, 11123–11129 (2003).
13. Okinaga, S. *et al.* C5L2, a nonsignaling C5A binding protein. *Biochemistry* **42**, 9406–9415 (2003).
14. Kalant, D. *et al.* C5L2 is a functional receptor for acylation stimulating protein. *J. Biol. Chem.* **280**, 23936–23944 (2005).
15. Riedemann, N. C. *et al.* Regulation by C5a of neutrophil activation during sepsis. *Immunity* **19**, 193–202 (2003).
16. Riedemann, N. C. *et al.* Regulatory role of C5a in LPS-induced IL-6 production by neutrophils during sepsis. *FASEB J.* **18**, 370–372 (2004).
17. Perianayagam, M. C., Balakrishnan, V. S., King, A. J., Pereira, B. J. & Jaber, B. L. C5a delays apoptosis of human neutrophils by a phosphatidylinositol 3-kinase-signaling pathway. *Kidney Int.* **61**, 456–463 (2002).
18. Hawlisch, H. *et al.* C5a negatively regulates Toll-like receptor 4-induced immune responses. *Immunity* **22**, 415–426 (2005).

19. Guo, R. F., Riedemann, N. C. & Ward, P. A. Role of C5a–C5aR interaction in sepsis. *Shock* **21**, 1–7 (2004).

20. Gavrilyuk, V. *et al.* Identification of complement 5a-like receptor (C5L2) from astrocytes: characterization of anti-inflammatory properties. *J. Neurochem.* **92**, 1140–1149 (2005).

21. Gao, H. *et al.* Evidence for a functional role of the second C5a receptor C5L2. *FASEB J.* **19**, 1003–1005 (2005).

22. White, J. H. *et al.* Heterodimerization is required for the formation of a functional GABA$_B$ receptor. *Nature* **396**, 679–682 (1998).

23. Sasaki, T. *et al.* Function of PI3Kγ in thymocyte development, T cell activation, and neutrophil migration. *Science* **287**, 1040–1046 (2000).

24. Duncan, G. S. *et al.* Genetic evidence for functional redundancy of Platelet/ Endothelial cell adhesion molecule-1 (PECAM-1): CD31-deficient mice reveal PECAM-1-dependent and PECAM-1-independent functions. *J. Immunol.* **162**, 3022–3030 (1999).

LETTERS

Endonuclease-independent LINE-1 retrotransposition at mammalian telomeres

Tammy A. Morrish[1]*†, José Luis Garcia-Perez[1]*, Thomas D. Stamato[3], Guillermo E. Taccioli[4], JoAnn Sekiguchi[1,2] & John V. Moran[1,2]

Long interspersed element-1 (LINE-1 or L1) elements are abundant, non-long-terminal-repeat (non-LTR) retrotransposons that comprise ~17% of human DNA[1]. The average human genome contains ~80–100 retrotransposition-competent L1s (ref. 2), and they mobilize by a process that uses both the L1 endonuclease and reverse transcriptase, termed target-site primed reverse transcription[3–5]. We have previously reported an efficient, endonuclease-independent L1 retrotransposition pathway (EN$_i$) in certain Chinese hamster ovary (CHO) cell lines that are defective in the non-homologous end-joining (NHEJ) pathway of DNA double-strand-break repair[6]. Here we have characterized EN$_i$ retrotransposition events generated in V3 CHO cells, which are deficient in DNA-dependent protein kinase catalytic subunit (DNA-PKcs) activity and have both dysfunctional telomeres and an NHEJ defect. Notably, ~30% of EN$_i$ retrotransposition events insert in an orientation-specific manner adjacent to a perfect telomere repeat (5'-TTAGGG-3'). Similar insertions were not detected among EN$_i$ retrotransposition events generated in controls or in XR-1 CHO cells deficient for XRCC4, an NHEJ factor that is required for DNA ligation but has no known function in telomere maintenance. Furthermore, transient expression of a dominant-negative allele of human *TRF2* (also called *TERF2*) in XRCC4-deficient XR-1 cells, which disrupts telomere capping, enables telomere-associated EN$_i$ retrotransposition events. These data indicate that L1s containing a disabled endonuclease can use dysfunctional telomeres as an integration substrate. The findings highlight similarities between the mechanism of EN$_i$ retrotransposition and the action of telomerase, because both processes can use a 3' OH for priming reverse transcription at either internal DNA lesions or chromosome ends[7,8]. Thus, we propose that EN$_i$ retrotransposition is an ancestral mechanism of RNA-mediated DNA repair associated with non-LTR retrotransposons that may have been used before the acquisition of an endonuclease domain.

We previously reported an endonuclease-independent pathway of L1 retrotransposition in a DNA-PKcs-deficient (V3) cell line (Fig. 1a; see also Supplementary Table 1)[6]. In addition to its role in NHEJ, the DNA-PKcs protein probably contributes to telomere capping, as embryonic fibroblasts derived from two *DNA-PKcs*$^{-/-}$ mouse strains exhibit an increase in anaphase bridges and telomere fusions[9,10]. Consistent with published results[10–12], V3 cells exhibit an increase in the number of anaphase bridges when compared to the AA8 parental cell line (Supplementary Fig. 1a). Moreover, 53BP1, a protein that is recruited to DNA double-strand breaks and dysfunctional telomeres, can localize to chromosomal ends in V3 cells[13–15] (Supplementary Fig. 1b, c).

To gain an insight into the mechanism of L1 retrotransposition in V3 cells, we characterized the sites of genomic integration. Four retrotransposition events generated from an endonuclease-proficient L1 showed the attributes of target-site primed reverse transcription (TPRT) because they are 5'-truncated, end in a poly(A) tail, integrate into an L1 endonuclease consensus cleavage site, and most were flanked by variable-length target-site duplications (Fig. 1b; insertions 1–4)[16,17]. The remaining insertion (Fig. 1b; insertion 5) is both 5'- and 3'-truncated and lacks a discernible target-site duplication. The 5' flank contains a 203-base-pair (bp) sequence that is homologous to a hamster sub-telomere-like repeat (AF306800; $P = 1 \times 10^{-41}$) that is also present at some pericentromeric and interstitial regions of CHO chromosomes[18]. Thus, most endonuclease-proficient L1 retrotransposition events in V3 cells occur by TPRT[6], although one insertion is associated with a telomere-like sequence at its 5' end and is similar structurally to previously described EN$_i$ retrotransposition events[6].

We next examined EN$_i$ retrotransposition events in V3 cells. Two insertions (Fig. 1b; insertions 8 and 9) were structurally similar to previously characterized EN$_i$ events (Fig. 1b; insertions 6 and 7)[6]. The last 532 bp flanking the 3' end of insertion 9 showed homology to the same hamster subtelomere-like repeat (AF306800 ($P = 7 \times 10^{-68}$)) flanking the 5' end of insertion 5. In comparison, the sequences flanking the 3' end of insertions 10 and 11 had different poly(A) tail lengths (~33 bases and ~120 bases, respectively) that are followed directly by the complement of a perfect series of telomere repeats (24 and 10 copies, respectively). The integrity of the repeats then degenerates and is followed by host DNA sequences. These data suggest that both EN$_i$ retrotransposition events initiated at related but distinct telomere repeats in V3 cells.

We next developed a polymerase chain reaction (PCR)/Southern blot assay to determine the frequency of telomere-associated EN$_i$ retrotransposition events (Fig. 2a). As controls, we amplified the L1–telomere junction fragments from genomic DNA of insertions 10 and 11 using a primer specific to the 3' end of our engineered L1 (that is, L1end326) and one of three primers specific for either the telomere–L1 junction (that is, telo10T and telo3T) or telomere sequence alone (that is, telo; see Methods). Southern blot analysis using an L1 probe (that is, L1end191) identified a ~326-bp product from insertion 10 genomic DNA using each primer set (Fig. 2b; lane 33-poly(A)). A heterogeneously sized product often was amplified from insertion 11 (Fig. 2b; lane 120-poly(A)), which probably reflects difficulty amplifying through the ~120-bp poly(A) tail. Sequencing confirmed the identity of the products (not shown).

The PCR assay was then used to screen genomic DNAs isolated from 17 EN$_i$ insertions generated in V3 cells. Experiments conducted with L1end326 and telo10T revealed products in cell lines 3L, 10E,

10L, 13E, 14E and 14L (Fig. 2b). In cell lines 10 and 14, the product was evident in both early (E) and a later (L) passage cell lines (see Methods). In cell line 3, we only detected a product in the later passage cell line, indicating that the insertion occurred during its establishment. In cell line 13, we only observed a product in the early passage cell line, suggesting a failed PCR reaction from the later passage cell line. We also observed faint bands of various sizes in other lanes (Fig. 2b, lanes 2E, 4L, 9E); however, sequencing revealed that only the intense bands in cell lines 3, 10, 13 and 14 contained bona fide L1–telomere junction fragments (see Fig. 2c).

To rule out nonspecific amplification owing to the ten deoxythymidine residues in the telo10T primer, PCR assays were conducted with the L1end326–telo3T or L1end326–telo primer pairs (Fig. 2b and data not shown). Again, cell lines 3, 10, 13 and 14 yielded positive signals, corroborating the above results. Sequencing revealed that the products have poly(A) tails that range from approximately 13–86 bases (Fig. 2c). However, we also detected a signal in cell line 5 (Fig. 2b, right panel). Sequencing demonstrated that the EN_i L1 insertion present in cell line 5 is truncated at the SV40 poly(A) cleavage site and only contains two deoxyadenosine residues at the L1–telomere junction (Fig. 2c), explaining why this product was not amplified using the L1end326–telo10T primer pair. Further analysis of the L1end326–telo PCR products revealed that the L1 poly(A) tail is always flanked by the complement of a perfect (5′-CCCTAA-3′) telomere repeat (Fig. 2c), and we confirmed these results using different telomere primers (Supplementary Fig. 2). Finally, primer pairs designed to detect telomere insertions in the opposite polarity did not yield PCR products (Supplementary Table 3a). These data indicate that the 3′ OH of a terminal telomere repeat (5′-TTAGGG-3′) was used to initiate reverse transcription.

To confirm EN_i retrotransposition at a chromosomal end, we performed fluorescence in situ hybridization (FISH) on V3 cells using a 4.5-kb probe containing host genomic DNA sequences flanking the telomere repeats present in insertion 10 (Figs 1b and 2d). If our engineered L1 inserted at a telomere end, this probe should hybridize to subtelomeric sequences. Approximately 83% of the positive FISH signals in cells at metaphase localized to the subtelomeric regions on five different chromosomes (Fig. 2d; see also Supplementary Table 2, categories A–D). The remaining cells at metaphase showed FISH signals located at different interstitial chromosomal locations (Supplementary Table 2, category E). We also observed di-centric chromosomes and karyotypic variation among individual V3 cells (Supplementary Table 2, category D), supporting the hypothesis that loss of the DNA-PKcs protein results in telomere dysfunction and genome instability[9,10,12].

We next examined whether an endonuclease-proficient L1 could retrotranspose at telomeres in parental AA8 cells and V3 cells (Supplementary Fig. 3e, f, and data not shown). As above, some faint bands were visible after Southern blotting; however, sequencing indicated that they were nonspecific products (data not shown).
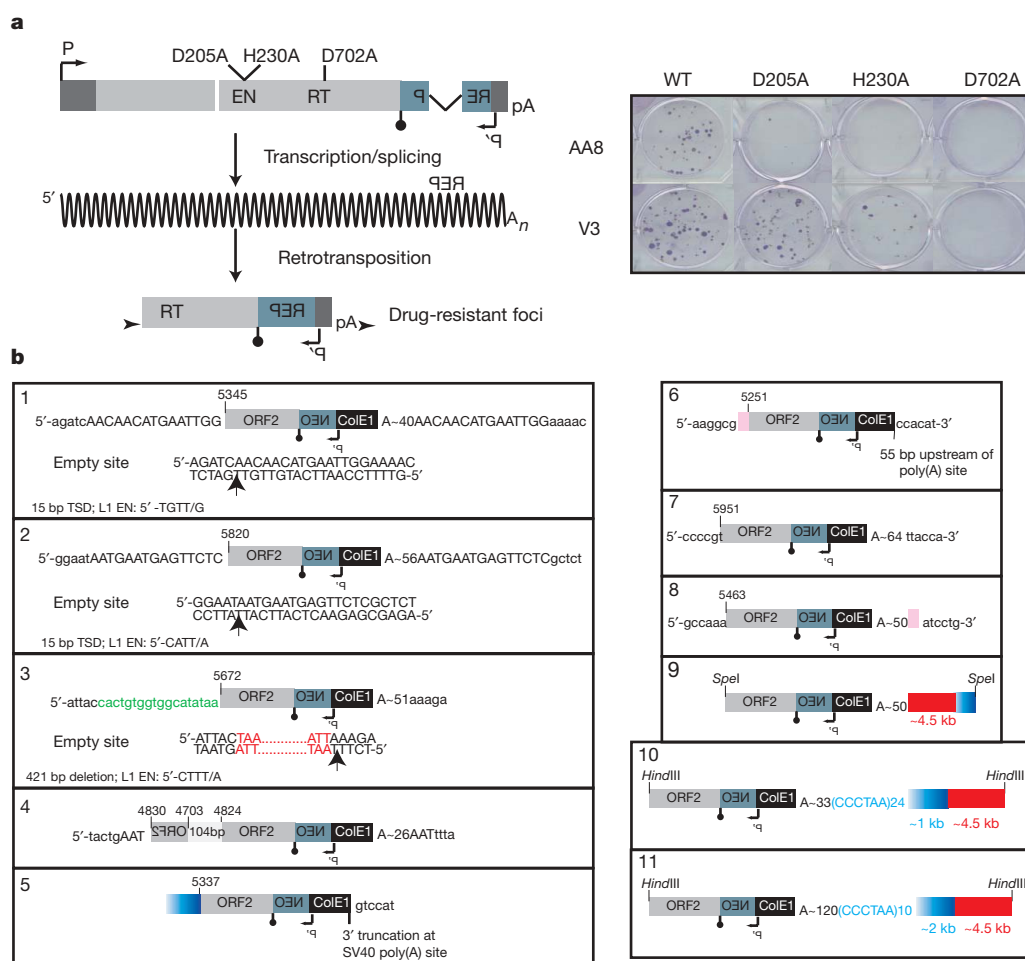

Figure 1 | Characterization of L1 retrotransposition insertions in V3 cells. a, A wild-type (WT) L1 with a retrotransposition indicator cassette (Rep). The EN (D205A and H230A) and RT (D702A) mutants are shown. Right: blasticidin-resistant foci in AA8 and V3 cells. **b**, Diagrams of wild-type (left) and EN_i (right) retrotransposition events. The pre-integration (insertions 1–3) and post-integration sites are shown. Arrowheads, L1 EN cleavage site; uppercase lettering, target site duplications (TSD); green lettering, non-L1 'filler' DNA[6]; blue lettering/rectangles, putative telomere/subtelomeric sequence; red and pink rectangles, genomic and cDNA sequences, respectively. L1 numbering is relative to L1.2 (GenBank accession M80343)[30]. (Additional details are provided in Supplementary Methods.)

209

Moreover, the faint products were not observed using primers that could amplify a larger L1–telomeric junction fragment (Supplementary Table 3). Thus, most wild-type L1 retrotransposition insertions in both AA8 and V3 cells seem to initiate by conventional TPRT[6], although it is possible that the sample size of our experiment is too small ($n = 49$) to detect telomere-associated events. Similar experiments failed to detect L1–telomere junction fragments among cohorts of EN_i retrotransposition events in either AA8 or XRCC4-deficient cells (Supplementary Figs 3f and 4), suggesting that telomere-associated retrotransposition is due to the dysfunctional telomere phenotype associated with V3 cells.

We then assayed EN_i retrotransposition in two clonal V3 cell lines that expressed a mouse DNA-PKcs complementary DNA (mDNA-PKcs comp5 and mDNA-PKcs comp15; Supplementary Fig. 3a). V3 cells exhibit increased radiation sensitivity and defects in a lymphoid-specific DNA rearrangement, V(D)J recombination, when compared with parental AA8 cells; the murine DNA-PKcs protein complements these phenotypes (Supplementary Fig. 3b, c). However, EN_i retrotransposition occurred at similar efficiencies in the V3 and DNA-PKcs complemented cell lines (Supplementary Table 1 and Supplementary Fig. 3d), and the complemented cells did not exhibit a significant decrease in anaphase bridges when compared to V3 cells (Supplementary Fig. 1a). Similar results were also obtained from V3 cell lines that were complemented with a human DNA-PKcs gene[19–23] (data not shown). Characterization of EN_i retrotransposition events
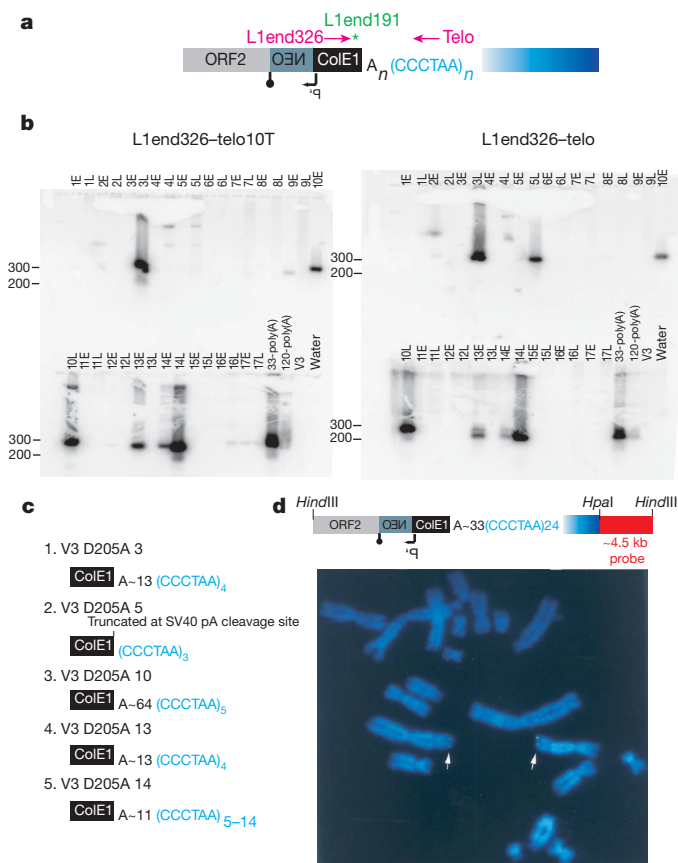

Figure 2 | EN_i retrotransposition initiates at the telomere. a, Primers used in the PCR (pink)/Southern blot (green) assay are shown above the diagram. **b,** Each lane contains PCR products from 17 DNAs from early (E) or later (L) passage clonal G418-resistant cell lines hybridized with L1end191. 33-poly(A) and 120-poly(A) are positive control DNAs from insertions 10 and 11 (Fig. 1b). Size markers (bp) and primer pairs are shown. **c,** Sequences of the PCR products using the L1end326–teloTTA primer pair. **d,** FISH results in V3 cells using the ~4.5-kb probe from insertion 10 (red rectangle). (Results are summarized in Supplementary Table 2.) White arrows indicate positive FISH signals.
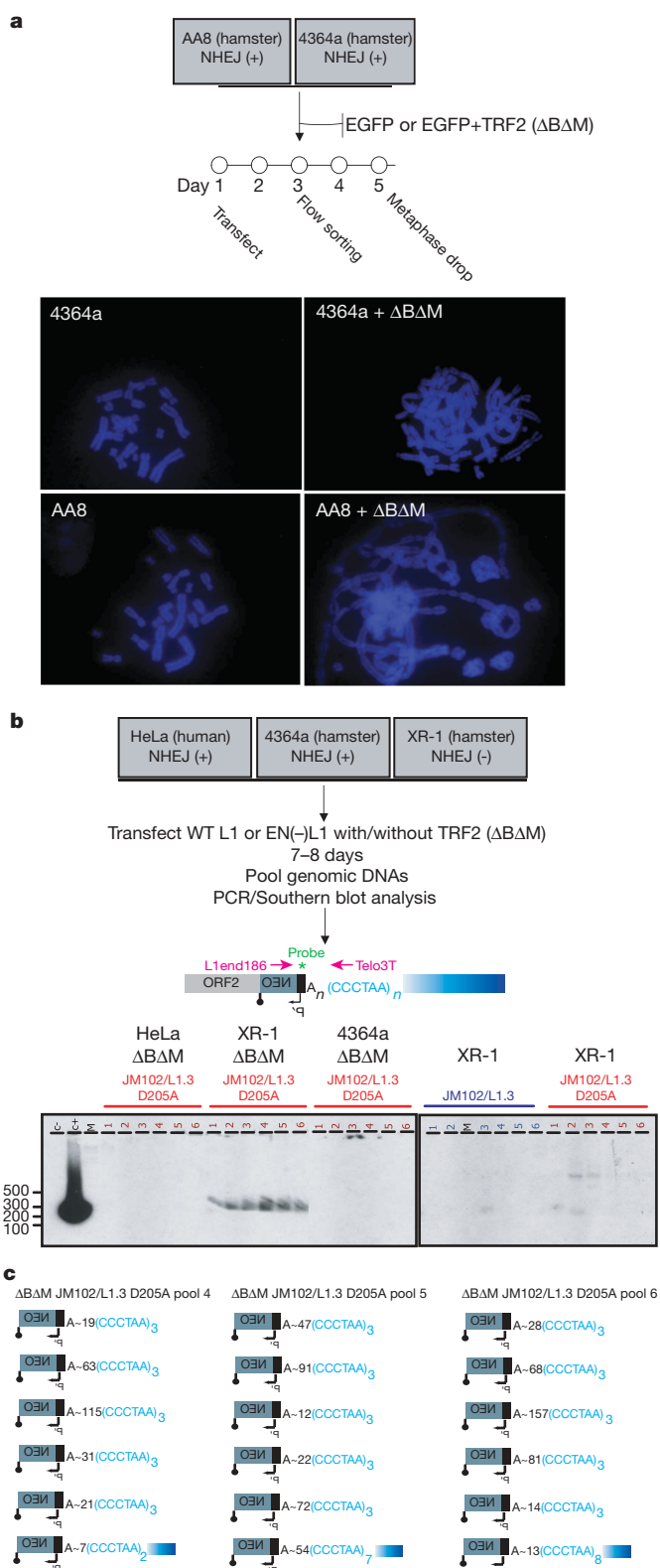

Figure 3 | Destabilization of telomeres in XRCC4-deficient cells allows for telomere-associated EN_i retrotransposition events. a, CHO cells transfected with an EGFP plasmid in either the presence or absence of a dominant-negative TRF2 ($\Delta B\Delta M$) expression plasmid were flow sorted and subjected to metaphase analysis. DAPI (4,6-diamidino-2-phenylindole)-stained metaphases from cells lacking (left panel) or containing (right panel) the TRF2 ($\Delta B\Delta M$) expression plasmid are shown. **b,** A wild-type or endonuclease-deficient L1 was transfected into the indicated cell line in either the presence or absence of the TRF2 ($\Delta B\Delta M$) plasmid and the resultant DNAs were subjected to the PCR/Southern blot assay. Size markers (bp) are shown. **c,** Sequences of the PCR products using the L1end186–telo3T primer pair.

in the complemented cell lines demonstrated an increase in telomere-associated events; moreover, we detected telomere-associated retrotransposition from a wild-type L1 (Supplementary Fig. 3e–g and Supplementary Table 3b).

The inability to complement the EN_i retrotransposition and anaphase bridge phenotypes in V3 cells requires further exploration. It is possible that V3 cells have accumulated lesions that are not present in the parental cell line, but contribute to the elevated efficiency of EN_i retrotransposition. Alternatively, species-specific differences between the DNA-PKcs proteins used in these experiments may prevent complementation of the dysfunctional telomere phenotype[20,24]. In either case, these data argue that the NHEJ defect and dysfunctional telomere phenotypes are genetically separable.

We then tested whether destabilizing telomeres in XRCC4-deficient cells by expressing a dominant-negative allele of human *TRF2* would allow a suitable substrate for EN_i retrotransposition[25]. TRF2 is a telomere-binding protein required for the protection of chromosome ends[15]. Consistent with previous studies[25,26], control experiments indicated that expression of the dominant-negative human TRF2 protein leads to chromosomal fusions in the NHEJ-proficient 4364a and AA8 parental cell lines (Fig. 3a). Next, we demonstrated that telomere-associated EN_i retrotransposition events are readily observed in XRCC4-deficient cells expressing the dominant-negative human TRF2 protein, but not in control cell lines expressing the same protein (Fig. 3b, c). These data are consistent with the hypothesis that a defect in NHEJ allows L1 to use a dysfunctional telomere as an integration substrate[26,27].

We conclude that EN_i retrotransposition can occur in an orientation-specific manner at terminal telomere repeats. We propose that dysfunctional telomeres generated either by a loss of DNA-PKcs or the expression of a dominant-negative allele of human *TRF2* in an NHEJ-deficient cell line allow the L1 retrotransposition machinery to use the 3′ OH present at a dysfunctional telomere as a primer to initiate EN_i retrotransposition (Fig. 4)[9–11,26,27]. The finding that the L1 reverse transcriptase does not require terminal base pairing between the primer and template to initiate reverse transcription provides biochemical support for this model[28].

We further speculate that integration of an L1 at the end of a chromosome would lead to chromosomal rearrangements, perhaps by initiating bridge-breakage-fusion cycles. Indeed the products of bridge-breakage-fusion cycles may provide recombination and/or

Figure 4 | Model for EN_i retrotransposition in NHEJ-deficient cells containing dysfunctional telomeres. A 3′ OH at a dysfunctional telomere can serve as a primer for reverse transcription of L1 RNA during EN_i retrotransposition. The blue rectangle represents the telomere of a chromosome; the green oval depicts the centromere; the orange line indicates L1 RNA; the red line indicates the L1 cDNA; the blue circle symbolizes the L1 RT. It is unclear whether base pairing between the L1 mRNA poly(A) tail and the thymidine residues at the telomere stabilizes priming, and how the resultant retrotransposition event is resolved. However, the event may be unstable, leading to secondary rearrangements and internal localization of the L1.

initiation substrates for EN_i retrotransposition, leading to the structural features observed in insertions 5 and 9, respectively (Fig. 1b). Notably, we cannot formally exclude the possibility that breaks at interstitial telomeres can also act as substrates for EN_i retrotransposition. However, the simplest explanation for our data is that dysfunctional telomeres can serve as substrates for EN_i retrotransposition.

METHODS

Oligonucleotide sequences and constructs used in this study. The sequences and constructs used in this study are described in detail in Supplementary Information.

L1 retrotransposition assay. The 3′ untranslated region of a retrotransposition-competent human L1 (L1.3) contains either a blasticidin or neomycin reporter (Rep) cassette designed to select retrotransposition events[4,6,16]. The relative positions of the CMV promoter (P) and SV40 late polyadenylation signal (pA) used for L1 expression are indicated in Fig. 1a. The relative positions of the SV40 promoter (P′) and the thymidine kinase polyadenylation signal (inverted 'lollipop') required for reporter gene expression are also shown. Some of the *mneoI*-based constructs include a prokaryotic promoter and a bacterial origin of replication (*ColE1*, black rectangle) to allow the recovery of retrotransposition events as autonomously replicating plasmids in *Escherichia coli*[16]. The reporter gene is interrupted by an intron in the same transcriptional orientation as the L1. This arrangement ensures that the reporter transcript will only be translated after L1 retrotransposition. A modified version of a transient transfection protocol was used to introduce the L1s into cells[6,29] (see Supplementary Information for details).

Recovery of retrotransposition events. Genomic DNA was isolated from G418-resistant clonal CHO cell lines harbouring an L1 retrotransposition event, and the events were characterized as described previously[16] (see Supplementary Information for details).

Analyses used. PCR and Southern blot analyses were performed using standard procedures (see Supplementary Information for details). FISH, anaphase bridge analyses, immunocytochemistry and metaphase analyses were performed using standard procedures (see Supplementary Information for details).

Characterization of mDNA-PKcs comp5 and comp15. The phenotypes of the mDNA-PKcs comp5 and mDNA-PKcs comp15 cell lines were characterized with respect to protein levels, V(D)J recombination, radiation sensitivity and EN_i retrotransposition (see Supplementary Information for details).

1. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Brouha, B. *et al.* Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl Acad. Sci. USA* **100**, 5280–5285 (2003).
3. Luan, D. D., Korman, M. H., Jakubczak, J. L. & Eickbush, T. H. Reverse transcription of R2Bm RNA is primed by a nick at the chromosomal target site: a mechanism for non-LTR retrotransposition. *Cell* **72**, 595–605 (1993).
4. Moran, J. V. *et al.* High frequency retrotransposition in cultured mammalian cells. *Cell* **87**, 917–927 (1996).
5. Feng, Q., Moran, J. V., Kazazian, H. H. Jr & Boeke, J. D. Human L1 retrotransposon encodes a conserved endonuclease required for retrotransposition. *Cell* **87**, 905–916 (1996).
6. Morrish, T. A. *et al.* DNA repair mediated by endonuclease-independent LINE-1 retrotransposition. *Nature Genet.* **31**, 159–165 (2002).
7. Pardue, M. L. & DeBaryshe, P. G. Retrotransposons provide an evolutionarily robust non-telomerase mechanism to maintain telomeres. *Annu. Rev. Genet.* **37**, 485–511 (2003).
8. Lingner, J. *et al.* Reverse transcriptase motifs in the catalytic subunit of telomerase. *Science* **276**, 561–567 (1997).
9. Goytisolo, F. A., Samper, E., Edmonson, S., Taccioli, G. E. & Blasco, M. A. The absence of the DNA-dependent protein kinase catalytic subunit in mice results in anaphase bridges and in increased telomeric fusions with normal telomere length and G-strand overhang. *Mol. Cell. Biol.* **21**, 3642–3651 (2001).
10. Gilley, D. *et al.* DNA-PKcs is critical for telomere capping. *Proc. Natl Acad. Sci. USA* **98**, 15084–15088 (2001).
11. Bailey, S. M. *et al.* DNA double-strand break repair proteins are required to cap the ends of mammalian chromosomes. *Proc. Natl Acad. Sci. USA* **96**, 14899–14904 (1999).
12. Bailey, S. M., Cornforth, M. N., Kurimasa, A., Chen, D. J. & Goodwin, E. H. Strand-specific postreplicative processing of mammalian telomeres. *Science* **293**, 2462–2465 (2001).
13. Takai, H., Smogorzewska, A. & de Lange, T. DNA damage foci at dysfunctional telomeres. *Curr. Biol.* **13**, 1549–1556 (2003).
14. Hao, L. Y., Strong, M. A. & Greider, C. W. Phosphorylation of H2AX at short telomeres in T cells and fibroblasts. *J. Biol. Chem.* **279**, 45148–45154 (2004).
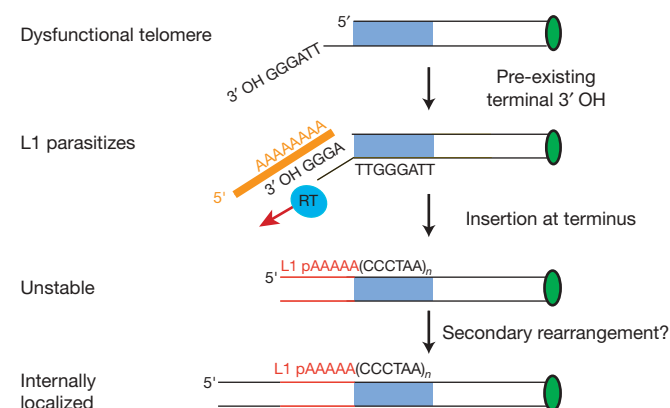
15. de Lange, T. Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev.* **19**, 2100–2110 (2005).

16. Gilbert, N., Lutz-Prigge, S. & Moran, J. V. Genomic deletions created upon LINE-1 retrotransposition. *Cell* **110**, 315–325 (2002).

17. Gilbert, N., Lutz, S., Morrish, T. A. & Moran, J. V. Multiple fates of l1 retrotransposition intermediates in cultured human cells. *Mol. Cell. Biol.* **25**, 7780–7795 (2005).

18. Faravelli, M. *et al.* Molecular organization of internal telomeric sequences in Chinese hamster chromosomes. *Gene* **283**, 11–16 (2002).

19. Taccioli, G. E. *et al.* Targeted disruption of the catalytic subunit of the DNA-PK gene in mice confers severe combined immunodeficiency and radiosensitivity. *Immunity* **9**, 355–366 (1998).

20. Blunt, T. *et al.* Identification of a nonsense mutation in the carboxyl-terminal region of DNA-dependent protein kinase catalytic subunit in the scid mouse. *Proc. Natl Acad. Sci. USA* **93**, 10285–10290 (1996).

21. Whitmore, G. F., Varghese, A. J. & Gulyas, S. Cell cycle responses of two X-ray sensitive mutants defective in DNA repair. *Int. J. Radiat. Biol.* **56**, 657–665 (1989).

22. Blunt, T. *et al.* Defective DNA-dependent protein kinase activity is linked to V(D)J recombination and DNA repair defects associated with the murine scid mutation. *Cell* **80**, 813–823 (1995).

23. Meek, K., Gupta, S., Ramsden, D. A. & Lees-Miller, S. P. The DNA-dependent protein kinase: the director at the end. *Immunol. Rev.* **200**, 132–141 (2004).

24. Beamish, H. J. *et al.* The C-terminal conserved domain of DNA-PKcs, missing in the SCID mouse, is required for kinase activity. *Nucleic Acids Res.* **28**, 1506–1513 (2000).

25. van Steensel, B., Smogorzewska, A. & de Lange, T. TRF2 protects human telomeres from end-to-end fusions. *Cell* **92**, 401–413 (1998).

26. Smogorzewska, A., Karlseder, J., Holtgreve-Grez, H., Jauch, A. & de Lange, T. DNA ligase IV-dependent NHEJ of deprotected mammalian telomeres in G1 and G2. *Curr. Biol.* **12**, 1635–1644 (2002).

27. Celli, G. B. & de Lange, T. DNA processing is not required for ATM-mediated telomere damage response after TRF2 deletion. *Nature Cell Biol.* **7**, 712–718 (2005).

28. Kulpa, D. A. & Moran, J. V. Cis-preferential LINE-1 reverse transcriptase activity in ribonucleoprotein particles. *Nature Struct. Mol. Biol.* **13**, 655–660 (2006).

29. Wei, W., Morrish, T. A., Alisch, R. S. & Moran, J. V. A transient assay reveals that cultured human cells can accommodate multiple LINE-1 retrotransposition events. *Anal. Biochem.* **284**, 435–438 (2000).

30. Dombroski, B. A., Mathias, S. L., Nanthakumar, E., Scott, A. F. & Kazazian, H. H. Jr. Isolation of an active human transposable element. *Science* **254**, 1805–1808 (1991).

Author Contributions T.A.M is co-first author and is a corresponding author. She contributed to the original concept, designed and performed experiments, analysed the data, and wrote the first draft of the manuscript. J.L.G.-P. is co-first author. He contributed to the concept, designed and performed experiments, analysed the data, and helped write and revise the manuscript. T.D.S and G.E.T. contributed reagents, helped with data interpretation and provided helpful comments during the course of this study. J.S. contributed the murine DNA-PKcs complemented cell lines, performed functional analyses on those cell lines, helped with data analysis, and helped revise the manuscript. J.V.M. is the senior author. He contributed to the original concept, analysed the data, revised the manuscript and provided financial support.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to J.V.M (moranj@umich.edu) or T.A.M. (morrisht@jhmi.edu).

Structure of an ABC transporter in complex with its binding protein

Kaspar Hollenstein[1], Dominik C. Frei[1] & Kaspar P. Locher[1]

ATP-binding cassette (ABC) transporter proteins carry diverse substrates across cell membranes[1]. Whereas clinically relevant ABC exporters are implicated in various diseases or cause multi-drug resistance of cancer cells[2,3], bacterial ABC importers are essential for the uptake of nutrients[4], including rare elements such as molybdenum. A detailed understanding of their mechanisms requires direct visualization at high resolution and in distinct conformations. Our recent structure of the multidrug ABC exporter Sav1866 has revealed an outward-facing conformation of the transmembrane domains coupled to a closed conformation of the nucleotide-binding domains, reflecting the ATP-bound state[5]. Here we present the 3.1 Å crystal structure of a putative molybdate transporter ($ModB_2C_2$) from *Archaeoglobus fulgidus* in complex with its binding protein (ModA). Twelve transmembrane helices of the ModB subunits provide an inward-facing conformation, with a closed gate near the external membrane boundary. The ATP-hydrolysing ModC subunits reveal a nucleotide-free, open conformation, whereas the attached binding protein aligns the substrate-binding cleft with the entrance to the presumed translocation pathway. Structural comparison of $ModB_2C_2A$ with Sav1866 suggests a common alternating access and release mechanism, with binding of ATP promoting an outward-facing conformation and dissociation of the hydrolysis products promoting an inward-facing conformation.

ABC transporters consist of two transmembrane domains (TMDs) that provide a translocation pathway, and two cytoplasmic nucleotide-binding domains (NBDs) that hydrolyse ATP and drive the transport reaction. The NBDs contain highly conserved motifs[6], whereas the TMDs have distinct primary sequences and even vary in the number of transmembrane helices. Although exporters feature a conserved core of 12 transmembrane helices, ABC importers contain between 10 and 20 transmembrane helices, depending on the mass and the chemical nature of the substrates they translocate. Also, the substrates of ABC exporters enter the translocation pathway from the cytoplasm or from the lipid bilayer, whereas ABC importers require a binding protein that captures the solutes and feeds them to the external face of the transporter.

Using X-ray crystallography, we have determined the structure of an archaeal molybdate ABC importer (*A. fulgidus* ModBC) in complex with its cognate binding protein (ModA). Although transport has not been demonstrated biochemically, the function of ModBC is suggested by the similarity of the primary sequences to bona fide molybdate transporters[7–9] and by trapping of a ModBC–A complex in the presence of ATP and orthovanadate (Supplementary Fig. 1), which is similar to observations with the bacterial maltose transporter[10]. The structure of $ModB_2C_2A$ (Fig. 1) was refined to 3.1 Å with good refinement statistics ($R_{work} = 25.6\%$; $R_{free} = 28.2\%$) and reveals a single ModA with bound substrate attached to the external side of $ModB_2C_2$. This arrangement allows unidirectional transport

by ABC importers, because only the binding proteins, not the TMDs, bind substrates effectively.

Each ModB subunit crosses the membrane 6 times for a total of 12 transmembrane segments in the transporter. The observed helix arrangement (Supplementary Fig. 3) is unprecedented and does not resemble that of the multidrug transporter Sav1866 (ref. 5) or the vitamin B_{12} importer BtuCD (ref. 11) or the putative metal chelate importer HI1470/71 from *Haemophilus influenzae*[12]. The ModB subunits form a large cavity, which is shielded from the lipid bilayer and
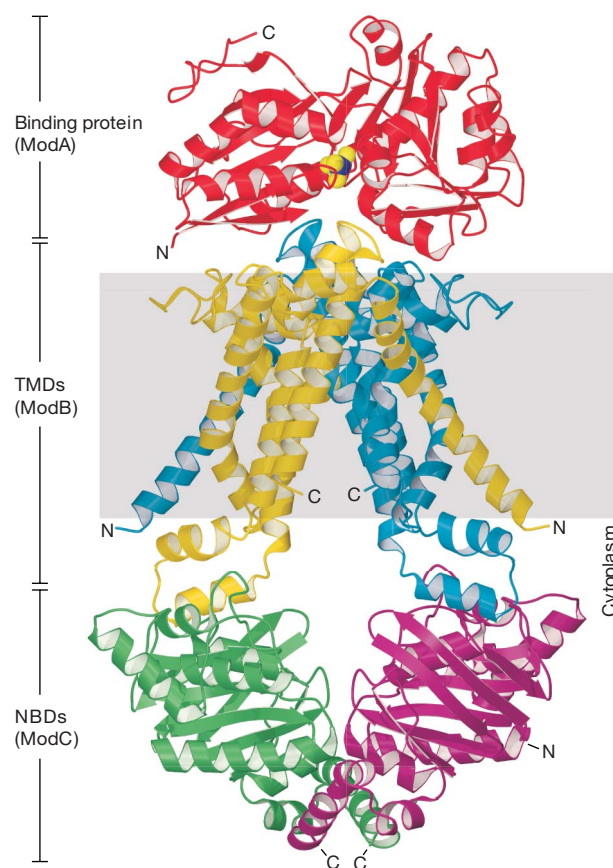

Figure 1 | Overall structure. Front view of the $ModB_2C_2A$ complex in ribbon representation, with the ModB subunits coloured yellow and blue, the ModC subunits coloured green and magenta, the binding protein ModA coloured red, and with bound tungstate in van der Waals representation (yellow and blue spheres). The grey box depicts the probable location of the lipid bilayer on the basis of the hydrophobicity of the protein surface. N, amino terminus; C, carboxy terminus. Note that there is a vertical two-fold molecular and non-crystallographic symmetry axis for $ModB_2C_2$.

[1]Institute of Molecular Biology and Biophysics, ETH Zurich, 8093 Zurich, Switzerland.

only accessible from the cytoplasm. The cavity, which represents the presumed translocation pathway, narrows towards the external membrane boundary and is separated from the extracytoplasmic space by a closed gate beneath the interface with the binding protein. The gate is formed by four converging protein stretches ('gate regions'), two from each ModB subunit (Fig. 2). Gate region 1 is located in transmembrane segment 3, whereas gate region 2 is located in transmembrane segment 5 and contains a conserved phenylalanine residue (Phe 200). The phenyl rings of Phe 200 from the ModB subunits are within van der Waals distance and their positioning suggests a role in gating. On aligning the primary sequence of *A. fulgidus* ModB with the TMDs of molybdate, sulphate, and phosphate importers, the gate regions are revealed as the most conserved motifs (Fig. 2c and Supplementary Fig. 4), suggesting a similar architecture.
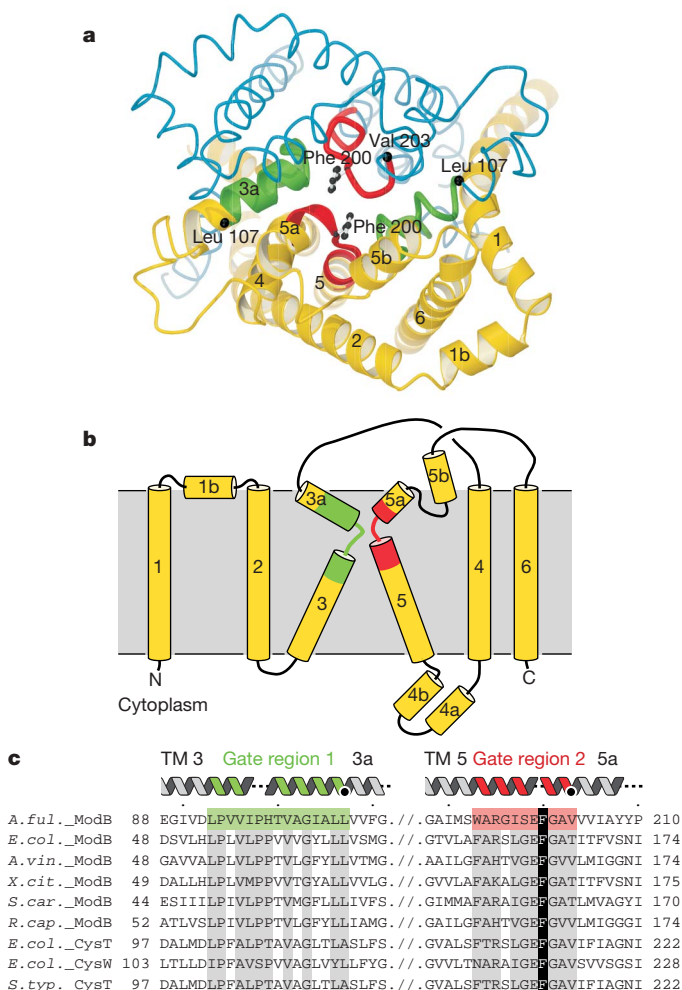
Figure 2 | ModB architecture and conserved gate regions. a, View from the extracellular side onto the transmembrane ModB subunits, with one subunit in yellow ribbon representation, the other as a blue backbone trace. Helices are numbered as in **b,** and conserved motifs forming the closed external gate are coloured green and red in both subunits. Relevant Cα positions are depicted as black spheres, with their residue numbers indicated. The conserved Phe 200 is shown in ball-and-stick representation. **b,** Topological scheme of the ModB subunit. Transmembrane helices are numbered consecutively, and short helices following a transmembrane helix additionally carry letters. The conserved gate regions are coloured as in **a. c,** Alignment of the *A. fulgidus* ModB sequence with those of other molybdate and sulphate ABC transporters (abbreviated species name is followed by an underscore and the ABC transporter name). Residues with significant conservation are shaded grey. Helices and conserved motifs are indicated above the sequences. Conserved gate regions are coloured as in **a.** TM, transmembrane.

In addition to the $ModB_2C_2A$ complex, we have determined the structures of isolated *A. fulgidus* ModA with bound molybdate and tungstate at 1.6 and 1.55 Å resolution, respectively (Supplementary Fig. 5 and Supplementary Table 2). Two lobes of ModA surround the substrate-binding site and are connected by a hinge that provides flexibility, allowing them to swing open in the absence of substrates, as observed in other binding proteins[13]. The main differences to the known structures of ModA from *Escherichia coli*[14] and *Azotobacter vinelandii*[15] include an additional, three-stranded β-sheet that is laterally appended, and, notably, an oxyanion binding site with octahedral coordination of the molybdenum and tungsten atoms, which is distinct from the tetrahedral coordination observed in *E. coli* and *A. vinelandii* ModA. The locations of structural change between ModA in isolation and when attached to the transporter are limited to the termini involved in lattice contacts and to the surface in contact with ModB where minor backbone shifts occur. ModA is bound to $ModB_2$ in an orientation that places the mouth of the substrate-binding cleft over the closed ModB gate that marks the entrance to the translocation pathway (Fig. 1). Both ModA lobes interact with ModB ($\sim$2,300 Å^2 buried surface area at the ModA–$ModB_2$ interface), in agreement with biochemical studies of the *Salmonella typhimurium* histidine transporter[16]. Several charged amino acid residues on the ModA surface contribute to the interface. The functional relevance of charged surface residues for the analogous interfaces has been demonstrated for the *S. typhimurium* histidine transporter and for the *Staphylococcus aureus* ferrichrome transporter, for which mutating such residues abolishes transport[17,18]. Two extracellular loops of ModB (one between transmembrane segments 3 and 4 and another between transmembrane segments 5 and 6) contribute to the interface with ModA. Because they predominantly interact with distinct ModA lobes, the conformations of these loops deviate from the symmetry that is otherwise present between ModB subunits.

The NBDs (ModC subunits) contain the highly conserved P-loops and LSGGQ motifs involved in ATP binding and hydrolysis[6]. The two ModC subunits reveal a conserved NBD fold[19] and are oriented in a 'head-to-tail' arrangement, with the P-loops of one subunit juxtaposed to the LSGGQ motif of the other. In the observed, nucleotide-free state, these motifs are separated by a gap, which has been referred to as the 'open' conformation[20]. In contrast, binding of ATP triggers a 'closed' conformation[20] as observed in the structures of isolated NBD dimers (for example, see refs 21–23) and Sav1866 (ref. 5). The observed spacing of the P-loops and the LSGGQ motifs in ModC is consistent with biophysical studies of the bacterial maltose transporter in which the solvent accessibility to the fluorescently labelled ATP binding sites is increased in the absence of ATP[24]. A comparison of the conformations of the NBDs of $ModB_2C_2A$ and Sav1866 is schematically depicted in Fig. 3a. The key differences in $ModC_2$ are: (1) a withdrawal of the helical subdomain from the interface with the opposite NBD, combined with a slight outward tilt; and (2) tilting of the full NBDs relative to one another, around a hinge at their carboxy termini. Combined, these changes result in a gap that faces the ModB subunits, with the C termini of ModC ensuring contact. The conformations of the nucleotide-free structures of BtuCD, MalK (ref. 23), HI1470/71 and ModC reveal remarkable and unexpected diversity. We conclude that although precise positioning of the P-loops and of the LSGGQ motifs is required in the ATP-bound state of ABC transporters, fewer geometrical constraints exist in the ATP-free state.

The ModC–ModB interface transmits critical conformational changes, thus coupling ATP binding and hydrolysis to transport. Primarily residues around the Q-loop[6] of ModC contribute to this interface, consistent with the structures of Sav1866, BtuCD and HI1470/71. The interface residues of ModB are located on the short cytoplasmic helices 4a and 4b, with 4a providing the bulk of the contacts. This helix ('coupling helix') interacts with a groove on the surface of ModC, at the boundary of the RecA-like subdomain and the helical subdomain (Fig. 3b). Similar coupling helices are

present in the structures of Sav1866, BtuCD and HI1470/71, and although the sequence similarities are limited[25], their striking superposition suggests a common role in coupling.

With the structures of four ABC transporters available, a mechanistic scheme emerges that rationalizes ATP-driven transport. The key to unidirectional transport seems to be the ATP-driven conversion of the inward-facing state, exemplified by $ModB_2C_2A$, to the outward-facing state, as visualized in Sav1866. Binding of two ATP molecules at the interface of the NBDs closes the gap between the conserved ATP-binding motifs. As this gap closes, so does the distance between the attached coupling helices, causing the TMDs to flip
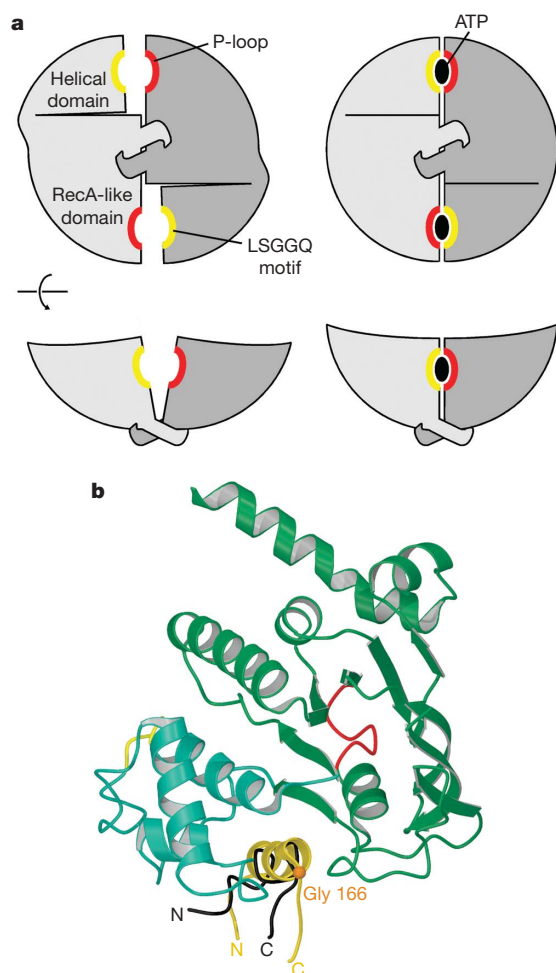

Figure 3 | NBD conformations and interfaces. a, Schematic summarizing the conformational changes observed by superimposing the NBDs of Sav1866 and *A. fulgidus* ModC. The light grey and dark grey areas represent individual NBDs, and the helical and the RecA-like subdomains are indicated. The upper panels represent a bottom view and the lower panels a side view of the two NBDs. The ATP-free state is depicted on the left, whereas the ATP-bound state is shown on the right. Red and yellow lines indicate the conserved P-loops (consensus sequence GxxGxGKST) and LSGGQ motifs. The hook-like features represent the C termini of the NBDs as observed in ModC, suggesting permanent contact of the NBDs throughout the reaction cycle. **b**, The architecturally conserved coupling helix of ABC transporters. A single ModC subunit in ribbon representation is shown with the RecA-like subdomains and helical subdomains coloured green and cyan, respectively. Segments corresponding to the P-loop and the LSGGQ motif are coloured red and light yellow, respectively. The 'coupling helix' (helix 4a) of ModB is indicated in dark yellow ribbon representation, with the Cα position of the conserved Gly 166 depicted as an orange sphere. After superposition of the NBDs, the analogous coupling helices of Sav1866, BtuCD and HI1470/71 superimpose well with that of ModB, but for clarity only that of Sav1866 is depicted in black. N and C reflect the amino and carboxy termini of the polypeptide stretches that form the coupling helices.

from the inward-facing to the outward-facing conformation. ABC exporters such as Sav1866 may then extrude bound drugs by presenting a cavity of low affinity to the extracellular medium[26]. In the ABC importer $ModB_2C_2A$, the rearrangements may not only affect the TMDs, but probably propagate to the attached binding protein, whose lobes they force apart, thus releasing bound substrate. The substrate may diffuse on a relatively straight trajectory from the binding site through the open gate and into the translocation pathway formed by the ModB subunits. The structure of BtuCD has revealed a cytoplasmic gate that was closed in the outward-facing conformation[11]. If a similar conformation exists in $ModB_2C_2$, molybdate may be trapped in the translocation pathway until hydrolysis of ATP and dissociation of the hydrolysis products allow the transporter to flip back to the inward-facing conformation. This mechanism can in principle explain uptake of nutrients by ABC importers and extrusion of cytotoxic drugs by ABC exporters[27]. It is consistent with the recently determined HI1470/71 structure, which revealed a nucleotide-free, inward-facing conformation. The vitamin B_{12} transporter BtuCD, in contrast, was observed in an outward-facing, nucleotide-free conformation. Although we cannot rule out a distinct mechanism for BtuCD, it seems more plausible that the purification and crystallization conditions may have trapped BtuCD in a conformation with the TMDs and NBDs partly uncoupled.

ABC transporters with two functional ATP binding sites hydrolyse two ATP molecules in each reaction cycle[28]. In $ModB_2C_2A$, a single binding protein is attached to the transporter, suggesting a stoichiometry of two ATP per imported molybdate, as was indeed demonstrated for the glycine/betaine ABC transporter OpuA from *Lactococcus lactis*[29]. With the structure of a complete ABC importer system at hand, kinetic studies[30] may now determine the sequence of events during the reaction cycle.

METHODS

A. fulgidus ModA and ModBC were separately overexpressed in *E. coli*. ModBC was solubilized and purified using octaethylene glycol monododecyl ether ($C_{12}E_8$) as a detergent, whereas ModA was purified in the absence of detergent. Purified $ModB_2C_2$ is active in detergent solution, as judged by its ATPase activity and characteristic inhibition by vanadate (Supplementary Fig. 1). Diffraction data of protein crystals were collected at the beam lines X06SA and X10SA of the Swiss Light Source (SLS). The $ModB_2C_2A$ structure was phased by multiple wavelength anomalous diffraction using bound tungstate in the ModA binding site. The experimental electron density was of high quality (Supplementary Fig. 2a) and model building was aided by the presence of selenomethionines serving as marker positions (Supplementary Fig. 2b). The structure of isolated ModA was phased using single wavelength anomalous diffraction from a selenomethionine derivative. Details of the biochemical and crystallographic procedures are given in the Supplementary Information. Buried surface areas were calculated with CNS (http://cns.csb.yale.edu), and figures were prepared using Molscript (www.avatar.se/molscript) and Raster3d (www.bmsc.washington.edu/raster3d).

1. Holland, I. B., Cole, S. P. C., Kuchler, K. & Higgins, C. F. *ABC Proteins: From Bacteria to Man* (Academic, London, 2003).
2. Gottesman, M. M. & Ambudkar, S. V. Overview: ABC transporters and human disease. *J. Bioenerg. Biomembr.* **33**, 453–458 (2001).
3. Gadsby, D. C., Vergani, P. & Csanady, L. The ABC protein turned chloride channel whose failure causes cystic fibrosis. *Nature* **440**, 477–483 (2006).
4. Nikaido, H. & Hall, J. A. Overview of bacterial ABC transporters. *Methods Enzymol.* **292**, 3–20 (1998).
5. Dawson, R. J. P. & Locher, K. P. Structure of a bacterial multidrug ABC transporter. *Nature* **443**, 180–185 (2006).
6. Schneider, E. & Hunke, S. ATP-binding-cassette (ABC) transport systems: Functional and structural aspects of the ATP-hydrolyzing subunits/domains. *FEMS Microbiol. Rev.* **22**, 1–20 (1998).
7. Rech, S., Deppenmeier, U. & Gunsalus, R. P. Regulation of the molybdate transport operon, modABCD, of *Escherichia coli* in response to molybdate availability. *J. Bacteriol.* **177**, 1023–1029 (1995).
8. Mouncey, N. J., Mitchenall, L. A. & Pau, R. N. Mutational analysis of genes of the *mod* locus involved in molybdenum transport, homeostasis, and processing in *Azotobacter vinelandii. J. Bacteriol.* **177**, 5294–5302 (1995).

9. Self, W. T., Grunden, A. M., Hasona, A. & Shanmugam, K. T. Molybdate transport. *Res. Microbiol.* **152,** 311–321 (2001).

10. Chen, J., Sharma, S., Quiocho, F. A. & Davidson, A. L. Trapping the transition state of an ATP-binding cassette transporter: Evidence for a concerted mechanism of maltose transport. *Proc. Natl Acad. Sci. USA* **98,** 1525–1530 (2001).

11. Locher, K. P., Lee, A. T. & Rees, D. C. The *E. coli* BtuCD structure: a framework for ABC transporter architecture and mechanism. *Science* **296,** 1091–1098 (2002).

12. Pinkett, H. W., Lee, A. T., Lum, P., Locher, K. P. & Rees, D. C. An inward-facing conformation of a putative metal-chelate-type ABC transporter. *Science* **315,** 373–377 (2007).

13. Quiocho, F. A. & Ledvina, P. S. Atomic structure and specificity of bacterial periplasmic receptors for active transport and chemotaxis: variation of common themes. *Mol. Microbiol.* **20,** 17–25 (1996).

14. Hu, Y., Rech, S., Gunsalus, R. P. & Rees, D. C. Crystal structure of the molybdate binding protein ModA. *Nature Struct. Biol.* **4,** 703–707 (1997).

15. Lawson, D. M., Williams, C. E., Mitchenall, L. A. & Pau, R. N. Ligand size is a major determinant of specificity in periplasmic oxyanion-binding proteins: the 1.2 Å resolution crystal structure of *Azotobacter vinelandii* ModA. *Structure* **6,** 1529–1539 (1998).

16. Liu, C. E., Liu, P. G., Wolf, A., Lin, E. & Ames, G. F. Both lobes of the soluble receptor of the periplasmic histidine permease, an ABC transporter (traffic ATPase), interact with the membrane-bound complex. Effect of different ligands and consequences for the mechanism of action. *J. Biol. Chem.* **274,** 739–747 (1999).

17. Prossnitz, E. Determination of a region of the HisJ binding protein involved in the recognition of the membrane complex of the histidine transport system of *Salmonella typhimurium*. *J. Biol. Chem.* **266,** 9673–9677 (1991).

18. Sebulsky, M. T., Shilton, B. H., Speziali, C. D. & Heinrichs, D. E. The role of FhuD2 in iron(III)-hydroxamate transport in *Staphylococcus aureus*. Demonstration that FhuD2 binds iron(III)-hydroxamates but with minimal conformational change and implication of mutations on transport. *J. Biol. Chem.* **278,** 49890–49900 (2003).

19. Jones, P. M. & George, A. M. The ABC transporter structure and mechanism: perspectives on recent research. *Cell. Mol. Life Sci.* **61,** 682–699 (2004).

20. Higgins, C. F. & Linton, K. J. The ATP switch model for ABC transporters. *Nature Struct. Mol. Biol.* **11,** 918–926 (2004).

21. Hopfner, K. P. *et al.* Structural biology of Rad50 ATPase: ATP-driven conformational control in DNA double-strand break repair and the ABC-ATPase superfamily. *Cell* **101,** 789–800 (2000).

22. Smith, P. C. *et al.* ATP binding to the motor domain from an ABC transporter drives formation of a nucleotide sandwich dimer. *Mol. Cell* **10,** 139–149 (2002).

23. Chen, J., Lu, G., Lin, J., Davidson, A. L. & Quiocho, F. A. A tweezers-like motion of the ATP-binding cassette dimer in an ABC transport cycle. *Mol. Cell* **12,** 651–661 (2003).

24. Mannering, D. E., Sharma, S. & Davidson, A. L. Demonstration of conformational changes associated with activation of the maltose transport complex. *J. Biol. Chem.* **276,** 12362–12368 (2001).

25. Mourez, M., Hofnung, N. & Dassa, E. Subunit interactions in ABC transporters: a conserved sequence in hydrophobic membrane proteins of periplasmic permeases defines an important site of interaction with the ATPase subunits. *EMBO J.* **16,** 3066–3077 (1997).

26. Loo, T. W. & Clarke, D. M. Recent progress in understanding the mechanism of P-glycoprotein-mediated drug efflux. *J. Membr. Biol.* **206,** 173–185 (2005).

27. Davidson, A. L. Mechanism of coupling of transport to hydrolysis in bacterial ATP-binding cassette transporters. *J. Bacteriol.* **184,** 1225–1233 (2002).

28. Sauna, Z. E. & Ambudkar, S. V. Evidence for a requirement for ATP hydrolysis at two distinct steps during a single turnover of the catalytic cycle of human P-glycoprotein. *Proc. Natl Acad. Sci. USA* **97,** 2515–2520 (2000).

29. Patzlaff, J. S., van der Heide, T. & Poolman, B. The ATP/Substrate stoichiometry of the ATP-binding cassette (ABC) transporter OpuA. *J. Biol. Chem.* **278,** 29546–29551 (2003).

30. Janas, E. *et al.* The ATP hydrolysis cycle of the nucleotide-binding domain of the mitochondrial ATP-binding cassette transporter Mdl1p. *J. Biol. Chem.* **278,** 26862–26869 (2003).

Tools of the trade

Even though the pace of drug discovery is hotting up, many candidate drugs fail late in development. Caitlin Smith looks at some of the tools used early in drug discovery that could help improve the situation.

The failure of hitherto promising drug candidates when they go into animal tests and clinical trials occurs far too often for the pharmaceutical industry's comfort. It represents costly failures for the companies involved and pushes up the overall cost of drug development.

Improving pipeline productivity is "clearly the biggest challenge", according to Kevin Hrusovsky, president and chief executive of microfluidics specialists Caliper Life Sciences in Hopkinton, Massachusetts.

Others agree. "The attrition rate of candidate drugs in later clinical development is still way too high," says Gary Franklin, industrial sector specialist in marketing communications at molecular-interactions company Biacore, based in Uppsala, Sweden, and recently acquired by GE Healthcare.

As well as interacting with its molecular target, a successful drug must also behave properly in the human body. It must be easily absorbed, not be broken down too quickly, and have no side effects that make it too toxic to use. Known as ADMET or ADMETox (for absorption, distribution, metabolism, excretion and toxicity), these tests are key to efficient drug discovery. ADMETox testing is initially carried out in

Kevin Hrusovsky: getting more drugs through the late stages of development is the big challenge.

CALIPER LIFE SCIENCES

cell-based and *in vitro* assays to determine properties such as a compound's solubility, its ability to cross cell membranes, and its cellular toxicity. But these tests are still far from perfect predictors, and improving them is a priority area in drug research. Undesirable ADMETox properties account for some 50–60% of drugs that fail at the preclinical stage.

Franklin believes that the biggest challenge is to obtain more comprehensive data much earlier in the drug-development process. He points to the Critical Path Initiative, published

in 2005 by the US Food and Drug Administration, which aims to stimulate a national effort to develop improved evaluation methods that provide better information and maximize the chances of clinical success.

To address this pressing need, researchers will need faster and more precise instruments that make measurements with greater information content than ever before. This is a big demand, but there are some new tools of the trade that may make a difference.

Automating ADMETox

Equipment companies have their sights set on improving ADMETox testing, including the tests that all new drugs must undergo for their possible adverse effects on the cardiac potassium channel (see 'Ion channels get automated').

Another routine ADMETox test predicts cellular uptake and efflux of a compound. The absorption of a compound across the epithelial lining of the human gut is conventionally predicted by assaying its transport into and out of Caco-2 cells. Caco-2 is a human colonic adenocarcinoma cell line that will differentiate in culture to express characteristic

ION CHANNELS GET AUTOMATED

Ion channels currently represent a less well charted territory of druggable targets than cell-surface receptors and enzymes, but this is changing. The electrophysiological techniques traditionally used to study ion-channel activity (one-cell, one-pipette patch clamping) are too slow for screening drug candidates, but some companies are finding ways to adapt and automate these techniques. One factor driving this development is the requirement by the regulatory authorities that all drugs must be tested for possible effects on a cardiac potassium channel (the hERG channel), which can cause cardiac arrhythmia. Compounds with adverse effects need to be weeded out as early as possible in the drug-discovery process.

The Dynaflow Pro II system from Cellectricon in Göteborg, Sweden, is based on a combination of patch clamping and microfluidics. "Dynaflow enables sequential
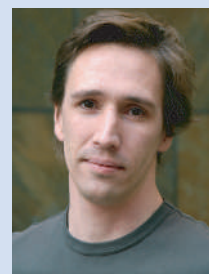
rather than parallel testing of compounds on ion channels by high-speed translational scanning of a single patch-clamped cell across a laminar stream of solution environments," says Mattias Karlsson, vice-president of research and development at Cellectricon.

Use of patch clamping retains the high quality of data characteristic of the technique, which is crucial in regulatory testing for the hERG channel, but the microfluidics component increases throughput compared with traditional lab methods.

For the highest-throughput automated patch-clamping, turn to Sophion Bioscience in Ballerup, Denmark. Sophion's QPatch HT is a 48-channel gigaseal patch-clamp system that can be used to study both voltage-gated and ligand-gated ion channels. Each channel, which patches one cell, is controlled individually, so the

system consists of 48 individual low-noise patch-clamp amplifiers and pressure controllers (for individual gigaseal formation). The liquid-handling robot controls eight pipettes. The recording chambers are contained within Sophion's QPlates, which house 48 glass-coated microfluidic channels that hold about 5 μl. With the small recording volumes and a liquid-exchange time of about 100 milliseconds, it is possible to screen multiple compounds on the same cell, or perform cumulative dose–response experiments.

Npi Electronic in Tamm, Germany, has incorporated a Tecan liquid-handling system into its ScreeningTool instrument for automated, fast and precise

Mattias Karlsson: applying microfluidics to patch-clamping.

CELLECTRICON

screening of ion-channel activity using two-electrode voltage clamping of *Xenopus* oocytes. Drugs are delivered by a rapid (millisecond resolution) automated system into the 15-ml bath of a miniature recording chamber. Npi plans to extend ScreeningTool to other cell types.

The ICR range of spectrometers from Aurora Biomed of Vancouver, British Columbia, provides a non-electrophysiological screen for ion-channel activity. The machines use atomic absorption spectroscopy and flux assays to detect activity and can be used to study both ligand- and voltage-gated channels. The new ICR 12000 is designed for ultra-high-throughput screening of compound libraries against ion-channel targets. C.S.

219

features of mature intestinal cells. The cells must be grown for 21 days in strictly controlled culture conditions to form a monolayer of differentiated cells before the test can be done. This tedious preparation process can be automated using a liquid-handling workstation such as the Biomek 3000, from Beckman Coulter of Fullerton, California, to perform the culture manipulations. "We have demonstrated intact and functional Caco-2 monolayers after 21 days of culture," says Keith Roby, product manager for strategic marketing automation at Beckman Coulter. "The workstation can also automate the compound permeability and efflux testing and sample collection."

The new Biomek FXP automated platform from Beckman Coulter can be used as a complete ADMETox workstation, and offers a faster *in vitro* assay for drug absorption than the Caco-2 assay. The new assay is based on parallel artificial membrane permeability analysis (PAMPA) and μSOL technologies for measurement of *in vitro* drug permeability and solubility, respectively, developed by pION of Woburn, Massachusetts.

In PAMPA, passive absorption through an artificial phospholipid membrane is intended to mimic the movement of a compound across the gut lining. "The PAMPA method provides significant time savings compared with the Caco-2 cell-based method, which takes more than three weeks to complete," says Roby. "With PAMPA, data can be collected in less than two hours," he says. All aspects of the permeability assay, including all liquid-handling steps and analysis, are controlled by PAMPA Evolution 96 software.

The Biomek FXP also supports metabolism assays, such as cytochrome P450 inhibition studies and microsomal stability assays. The toxicology assays it supports include ones based on cell proliferation, cytotoxicity and apoptosis.

The small-footprint Biomek NX Assay Workstation can also be used to automate cell-based screens for apoptosis. The system will schedule a six-hour method to process ten assay plates from the initial induction of the apoptosis pathway to detection of activity of the effector enzyme caspase 3/7, without the intervention of an operator.

Mark Collins, associate director of strategic marketing at Thermo Fisher Scientific, based in Waltham, Massachusetts, also considers late-stage failure of drug candidates a big problem for the pharmaceutical industry. "The attrition rate is one of the biggest issues; too many compounds fail at preclinical or phase I clinical tests," he says.

Collins thinks that high-content screening (HCS) could be one of the answers. "ADMETox testing is starting to use HCS for automated

Look, no cells: Beckman Coulter's Biomek FX ADMETox Workstation can be used with pION's PAMPA technology for permeability assays using synthetic phospholipid membranes.

genotoxicity assays as well as for predictive assays to rank compounds for animal testing. This research may pave the way for HCS assays to be a better predictor of certain types of toxicity than animals," he says.

On the software front, both Inpharmatica of London, UK, (soon to be part of Belgian drug-discovery company Galapagos) and ChemSilico of Tewksbury, Massachusetts, have software tools to predict ADME properties.

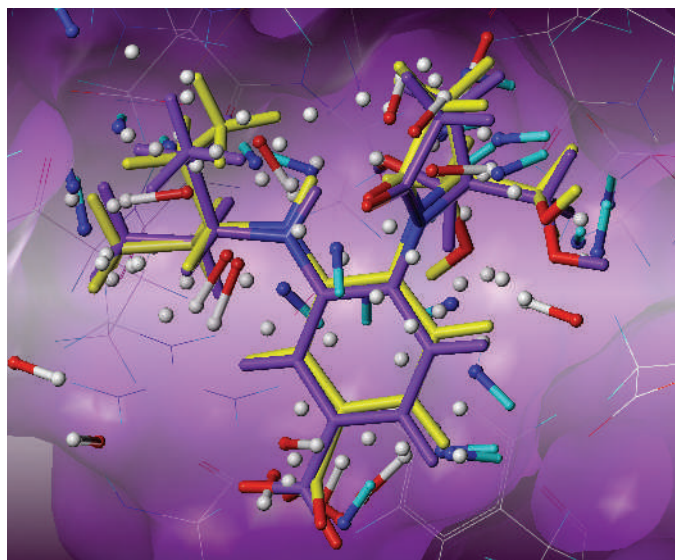## Intelligent cell screening

"The continued growth of cell-based assays is driving a need for automated systems to prepare and run these assays," says Marc Feiglin, chief technology officer in life-science research at Tecan in Männedorf, Switzerland. "Automating this in a cost-effective manner requires a single system that can pipette both large volumes of media and small volumes of compounds."

Model virus: a potential influenza virus inhibitor (purple) modelled with Tripos's SYBYL software and redocked to the viral neuraminidase (yellow).

The Freedom EVO platform from Tecan offers versatile robotic liquid-handling. It is available in four sizes, with a variety of robotic arms and other liquid-handling accessories. According to Feiglin, it offers the widest dynamic range of pipetting volumes currently available on one platform — from less than 1 μl to hundreds of millilitres.

The Cellomics family of automated imaging instruments from Thermo Fisher Scientific are designed for HCS and high-content analysis. They enable researchers to analyse the responses of cells *in vitro* to stimulation by small molecules or the knockdown of gene expression by RNA interference (RNAi). "HCS provides temporal and spatial information about such cellular responses as well as phenotypic profiles," says Collins. The core of a typical Cellomics HCS system is an imager such as their ArrayScan VTI. This is packaged with quantitative image-analysis software modules that address the experimental area of interest and data-management software, necessary because of the huge quantities of imaging data that HCS generates.

A feature of the Cellomics HCS system is intelligentAcQuisition (iQ), which allows images to be analysed on-the-fly in real time, and measurements from the analysis fed back to allow the image-acquisition process to adapt to the biology being measured. "IQ provides significant productivity gains compared with HCS systems that acquire images and then do the analysis offline, as it performs in parallel all the steps of the HCS process during the time the plate is being scanned," says Collins. He likens Cellomics iQ to a digital camera, which allows you to see immedi-

TRIPOS

ate results. Without it you're more like someone taking pictures with a film camera: "You don't know how good the picture is until you have the film developed, but by then it might be too late, or you have too many pictures of the same thing," he says.

Picking up the signal

One application of Cellomics HCS systems is to study intracellular signalling pathways in oncology, as they enable a pathway to be 'walked' using multiplexing.

Another company pursuing ways to streamline signalling research is DiscoveRx of Fremont, California, which received a 2006 Frost & Sullivan award for technology innovation for its PathHunter assay platform, which is used to study signalling pathways in intact cells. A recent addition to the range is a β-arrestin assay to study the activation of G-protein-coupled receptors (GPCRs), a common drug target.

An attractive feature of PathHunter assays is that they are relatively simple: they are adapted to microtitre plates, involve only one or two additions of reagents, and do not require cell washing or fixation. The assays are compatible with many different cell lines, especially those most commonly used in high-throughput screening, such as CHO and HEK cells.

PathHunter assays detect receptor activation by a technique called enzyme fragment complementation. Two modified β-galactoside fragments, fused to the receptor and the β-arrestin molecule, respectively, are brought together when β-arrestin binds to the activated receptor and form an active enzyme that generates a chemiluminescent signal.

DiscoveRx claims that PathHunter is the first chemiluminescence assay technology that can measure protein trafficking directly inside the cell — as such, it is especially suited to evaluating how cellular physiology is perturbed by drug candidates. It can be used to study translocation, degradation, secretion, protein–protein interactions and membrane trafficking. The chemiluminescent signal generated does not need dedicated imaging technology and can be read in 96-, 384- or 1,536-well microplates with a standard luminometer.

Tecan's new Infinite series of microplate readers are compatible with its Freedom EVO liquid-handling workstation and are available as either filter- or monochrometer-based systems. According to Feiglin, the sensitivity of the luminescence readers makes them a popular choice for assays of GPCR activity. Using the 1,536-well optics options on the Infinite F500 filter-based microplate reader, a 1,536-well plate can be read in less than 30 seconds. Also new from Tecan is the HydroFlex automated microplate washing and vacuum filtration system, which can be integrated into the Freedom EVO platform.

CyBio of Jena, Germany, made its reputation as one of the first companies to offer multipipettors. It recently unveiled a new robotic liquid-handling platform, CyBi-RoboSpense, which will work with almost any type of sample tube or plate, and can even transport plates to other instruments such as

Tecan's HydroFlex platewasher can be integrated into the Freedom EVO system.

other pipettors, microplate readers or liquid chromatography or mass spectroscopy equipment. The CyBi-RoboSpense can also be used for complicated sample preparation and purification routines such as tryptic digests and magnetic bead separation.

Microfluidics applications

More companies are incorporating the principles of microfluidics into their instruments, making the movement of microlitre — or even nanolitre — volumes of solutions do what used to take hundreds of millilitres, not to mention a saving in time and labour. An innovator in this area is Gyros of Uppsala, Sweden, whose Gyrolab Bioaffy microlabs for protein purification and quantification are based on affinity-capture techniques. The Gyrolab Bioaffy 200 is ideal for high-sensitivity assays to detect low concentrations of proteins. It can be used, for example,

IN SILICO SCREENING WITH CHEMICAL INFORMATICS

Imagine evaluating the efficacy and safety of new drugs without having to lift a pipette. Advances in *in silico* drug development are beginning to make this possible.
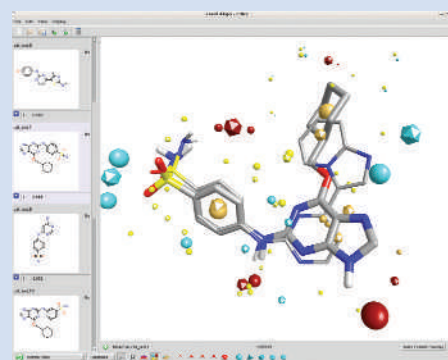
The SYBYL range of computational chemistry software from Tripos of St Louis, Missouri, will design compound libraries, select the best binding partners and screen potential drug molecules for safety and efficacy on the basis of known protein target and/or ligand structures — all *in silico*. A new advanced protein-modelling module predicts the structure of proteins of unknown structure using related proteins of known structure as the models. "It rapidly and accurately provides a valuable protein structure to be used for *in silico* docking experiments," says Kathleen Mensler, vice-president of marketing and corporate development at Tripos. Another new product, Surflex-Sim, helps

find new classes of compounds that are structurally similar to already known, suboptimal, lead compounds, but that may be more active or safer.

Modelling software from Cresset BioMolecular Discovery in Letchworth, UK, uses surface properties or 'molecular fields' of proteins and other molecules to predict intermolecular interactions. "We know that structurally diverse compounds can bind to the same protein site and elicit the same response, but this has never been fully explained by molecular structure," says Sally Rose, Cresset's director of business development. "Our view is that molecular fields provide vital information for understanding and predicting activity," she says. The fields are derived by calculating the interaction energy of a 'probe' atom carrying a positive, negative or neutral charge with the compound. "Our fields summarize the key

binding information, and we locate the most important regions around a ligand where the fields are strongest and binding is expected," says Rose. The original FieldScreen software for virtual screening has been joined by FieldTemplater, designed to find three-dimensional bioactive conformations, and FieldAlign, a molecular-alignment tool for comparing the three-dimensional structures of two molecules.

A pioneer of *in silico* modelling, De Novo Pharmaceuticals in Cambridge, UK, recently announced a new focus for its drug-discovery programme — finding drug targets in various metabolic diseases. De Novo has upgraded SkelGen, its proprietary

Line up: Cresset's FieldAlign program compares molecular structures.

in silico drug-design platform, with the addition of Reflex, which takes into account the rotational flexibility of amino acids in the target's active site. This better evaluates the shape of the active site and its interaction with small-molecule ligands, so that the modelling more closely resembles the dynamics of real protein–ligand interactions.   C.S.

221

to quantify protein markers of disease present in the nanograms per millilitre range. The Gyrolab Bioaffy 20HC is designed for bio-pharmaceutical applications such as process optimization and the quantification of monoclonal antibody drugs intended for therapeutic use, especially for measurements on the milligrams per millilitre scale, which means that samples can be run without dilution.

The CD-sized microlabs run on Gyrolab workstations; the Gyrolab Workstation LIF can run up to five CDs per batch, and includes a laser-induced fluorescence (LIF) detector. "What made this system particularly interesting to the customer," says Christina Burtsoff Asp, vice-president of marketing at Gyros, "was that it provided a measurement range from micrograms to grams per litre, which allowed dilution-free analysis in concentration ranges relevant for the entire purification process, from cell supernatant to purified product — all on a single system."

Another new product using microfluidic technology is Caliper's Desktop Profiler, a bench-top kinase-profiling system that is pre-loaded with microplates with protein kinases, substrates and reagents. "Kinase profiling enables the identification of potential side effects in therapeutic candidates, thus helping researchers to better understand how a compound will react in the human body," explains Hrusovsky. The Desktop Profiler uses Caliper's LabChip microfluidics-based screening technology, and is designed for researchers seeking to qualify drug candidates — that is, to select the more promising drug candidates from those that might be too toxic or ineffectual.

Fluidigm in San Francisco, California, recently introduced a new digital array
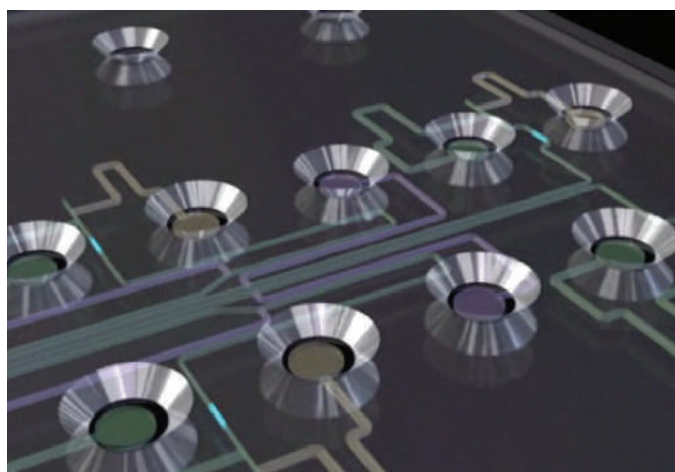
Micromanagement: a computer rendition of the surface of a Caliper LabChip shows the circular wells through which the channels are loaded.

nanofluidic chip for a type of quantitative PCR that the company calls digital PCR, which runs on its BioMark system. Digital PCR is based on limiting-dilution analysis, and estimates the absolute number of copies of, for example, a target sequence in a sample by successively diluting the sample and partitioning it into a large number of chambers, such that each chamber contains either one or no target molecules.

Nanofluidics makes this application of limiting-dilution analysis practicable, in that after the sample is loaded, it is automatically distributed into about 1,000 nanolitre-volume reaction chambers. Amplification of the target sequence by PCR in chambers that contain a molecule can be detected and the number of original molecules in the sample calculated.

The new technology is already proving useful: one group of researchers has used it to correlate specific copies of genes with the regulatory states of single stem cells; another study used digital arrays to identify individual bacterial cells carrying marker genes. This technology promises to compete with the microwell plates currently used for high-throughput quantitative PCR. Digital arrays offer higher throughput, use smaller amounts of reaction solutions, and avoid the thousands of pipetting steps that would be needed for a similar experiment in microwell plates.

Going *in vivo*

However sophisticated cell-based and *in vitro* tests become, promising candidate drugs have eventually to be tested in living organisms. Hrusovsky points to the problem of the poor correlation of *in vitro* experiments with what happens *in vivo*. "Pipeline productivity

can be improved by preventing the drug attrition that happens when a candidate compound moves from *in vitro* to *in vivo* studies," he says.

Caliper's IVIS Spectrum *in vivo* optical-imaging system uses tomography to give three-dimensional images of fluorescent and bioluminescent reporters in whole organisms. It is designed to help researchers "better visualize, track and understand how a therapeutic compound reacts, at a molecular level, *in vivo*, thus bridging *in vitro* and *in vivo* experimentation", says Hrusovsky. Spectral unmixing capabilities allow multiple fluorescent reporters to be used. Illumination can be switched from above the specimen to below it, enabling the tomographic reconstruction of, for example, both shallow and deep tumours.

Getting more selective

Several general trends may speed up the drug-discovery process and reduce the rate of attrition of drug candidates. Franklin notes a shift towards structure-based techniques for drug discovery and the identification of new druggable targets (see '*In silico* screening with chemical informatics'), using dramatically smaller compound libraries compared with high-throughput screening. "This has the potential to significantly shorten drug-discovery timelines and, hopefully, to increase overall productivity in the industry," he says. "The trend towards fragment-based screening in particular shows great promise."

"More rational screening — that is, for higher information content — of smaller libraries would help," agrees Collins. "Too many targets from molecular biology and genomics have little or no clinical relevance; high-content screening and other technologies such as RNAi can help here." Collins also sees pressures on the drug-discovery industry to produce 'personalized medicine', which would provide a drug that best fits the individual patient. Traditionally, the drug industry "has sought the one-size-fits-all blockbuster, so to respond to the idea of personalized medicine, technological, economic, and ethical changes will be needed", he says.

"Some believe that the industry is poised to see the benefits of the investments made in new technologies over the past few years, and that state-of-the-art informatics is finally capable of supporting the complexity of life-science discovery," says Kathleen Mensler, vice-president of marketing and corporate development at modelling software company Tripos in St Louis, Missouri. "With advances in understanding biological pathways and the unprecedented amount of valuable data available, breakthroughs are inevitable. The question is when." ■

Caitlin Smith is a freelance writer in Portland, Oregon.

TRIPOS
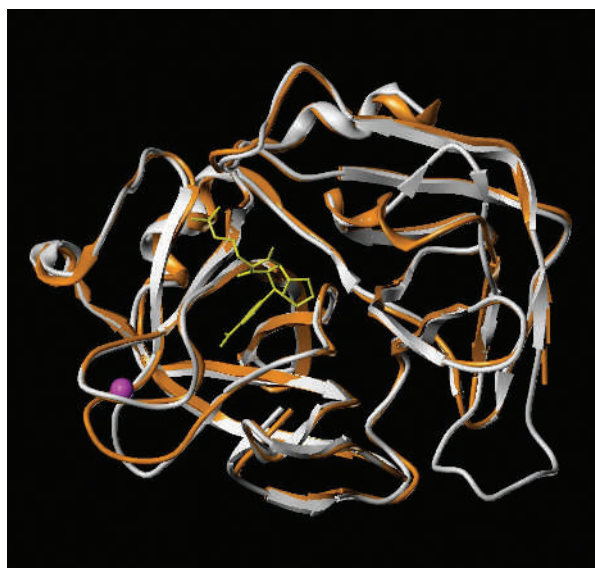
Structure from sequence: a homology model (white) can match a crystal structure (orange) remarkably well.

Tools of the trade

Even though the pace of drug discovery is hotting up, many candidate drugs fail late in development. Caitlin Smith looks at some of the tools used early in drug discovery that could help improve the situation.

The failure of hitherto promising drug candidates when they go into animal tests and clinical trials occurs far too often for the pharmaceutical industry's comfort. It represents costly failures for the companies involved and pushes up the overall cost of drug development.

Improving pipeline productivity is "clearly the biggest challenge", according to Kevin Hrusovsky, president and chief executive of microfluidics specialists Caliper Life Sciences in Hopkinton, Massachusetts.

Others agree. "The attrition rate of candidate drugs in later clinical development is still way too high," says Gary Franklin, industrial sector specialist in marketing communications at molecular-interactions company Biacore, based in Uppsala, Sweden, and recently acquired by GE Healthcare.

As well as interacting with its molecular target, a successful drug must also behave properly in the human body. It must be easily absorbed, not be broken down too quickly, and have no side effects that make it too toxic to use. Known as ADMET or ADMETox (for absorption, distribution, metabolism, excretion and toxicity), these tests are key to efficient drug discovery. ADMETox testing is initially carried out in

Kevin Hrusovsky: getting more drugs through the late stages of development is the big challenge.

CALIPER LIFE SCIENCES

cell-based and *in vitro* assays to determine properties such as a compound's solubility, its ability to cross cell membranes, and its cellular toxicity. But these tests are still far from perfect predictors, and improving them is a priority area in drug research. Undesirable ADMETox properties account for some 50–60% of drugs that fail at the preclinical stage.

Franklin believes that the biggest challenge is to obtain more comprehensive data much earlier in the drug-development process. He points to the Critical Path Initiative, published

in 2005 by the US Food and Drug Administration, which aims to stimulate a national effort to develop improved evaluation methods that provide better information and maximize the chances of clinical success.

To address this pressing need, researchers will need faster and more precise instruments that make measurements with greater information content than ever before. This is a big demand, but there are some new tools of the trade that may make a difference.

Automating ADMETox

Equipment companies have their sights set on improving ADMETox testing, including the tests that all new drugs must undergo for their possible adverse effects on the cardiac potassium channel (see 'Ion channels get automated').

Another routine ADMETox test predicts cellular uptake and efflux of a compound. The absorption of a compound across the epithelial lining of the human gut is conventionally predicted by assaying its transport into and out of Caco-2 cells. Caco-2 is a human colonic adenocarcinoma cell line that will differentiate in culture to express characteristic

ION CHANNELS GET AUTOMATED

Ion channels currently represent a less well charted territory of druggable targets than cell-surface receptors and enzymes, but this is changing. The electrophysiological techniques traditionally used to study ion-channel activity (one-cell, one-pipette patch clamping) are too slow for screening drug candidates, but some companies are finding ways to adapt and automate these techniques. One factor driving this development is the requirement by the regulatory authorities that all drugs must be tested for possible effects on a cardiac potassium channel (the hERG channel), which can cause cardiac arrhythmia. Compounds with adverse effects need to be weeded out as early as possible in the drug-discovery process.

The Dynaflow Pro II system from Cellectricon in Göteborg, Sweden, is based on a combination of patch clamping and microfluidics. "Dynaflow enables sequential
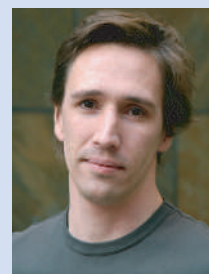
rather than parallel testing of compounds on ion channels by high-speed translational scanning of a single patch-clamped cell across a laminar stream of solution environments," says Mattias Karlsson, vice-president of research and development at Cellectricon.

Use of patch clamping retains the high quality of data characteristic of the technique, which is crucial in regulatory testing for the hERG channel, but the microfluidics component increases throughput compared with traditional lab methods.

For the highest-throughput automated patch-clamping, turn to Sophion Bioscience in Ballerup, Denmark. Sophion's QPatch HT is a 48-channel gigaseal patch-clamp system that can be used to study both voltage-gated and ligand-gated ion channels. Each channel, which patches one cell, is controlled individually, so the

system consists of 48 individual low-noise patch-clamp amplifiers and pressure controllers (for individual gigaseal formation). The liquid-handling robot controls eight pipettes. The recording chambers are contained within Sophion's QPlates, which house 48 glass-coated microfluidic channels that hold about 5 μl. With the small recording volumes and a liquid-exchange time of about 100 milliseconds, it is possible to screen multiple compounds on the same cell, or perform cumulative dose–response experiments.

Npi Electronic in Tamm, Germany, has incorporated a Tecan liquid-handling system into its ScreeningTool instrument for automated, fast and precise

Mattias Karlsson: applying microfluidics to patch-clamping.

CELLECTRICON

screening of ion-channel activity using two-electrode voltage clamping of *Xenopus* oocytes. Drugs are delivered by a rapid (millisecond resolution) automated system into the 15-ml bath of a miniature recording chamber. Npi plans to extend ScreeningTool to other cell types.

The ICR range of spectrometers from Aurora Biomed of Vancouver, British Columbia, provides a non-electrophysiological screen for ion-channel activity. The machines use atomic absorption spectroscopy and flux assays to detect activity and can be used to study both ligand- and voltage-gated channels. The new ICR 12000 is designed for ultra-high-throughput screening of compound libraries against ion-channel targets.   C.S.

ate results. Without it you're more like someone taking pictures with a film camera: "You don't know how good the picture is until you have the film developed, but by then it might be too late, or you have too many pictures of the same thing," he says.

Picking up the signal

One application of Cellomics HCS systems is to study intracellular signalling pathways in oncology, as they enable a pathway to be 'walked' using multiplexing.

Another company pursuing ways to streamline signalling research is DiscoveRx of Fremont, California, which received a 2006 Frost & Sullivan award for technology innovation for its PathHunter assay platform, which is used to study signalling pathways in intact cells. A recent addition to the range is a β-arrestin assay to study the activation of G-protein-coupled receptors (GPCRs), a common drug target.

An attractive feature of PathHunter assays is that they are relatively simple: they are adapted to microtitre plates, involve only one or two additions of reagents, and do not require cell washing or fixation. The assays are compatible with many different cell lines, especially those most commonly used in high-throughput screening, such as CHO and HEK cells.

PathHunter assays detect receptor activation by a technique called enzyme fragment complementation. Two modified β-galactoside fragments, fused to the receptor and the β-arrestin molecule, respectively, are brought together when β-arrestin binds to the activated receptor and form an active enzyme that generates a chemiluminescent signal.

DiscoveRx claims that PathHunter is the first chemiluminescence assay technology that can measure protein trafficking directly inside the cell — as such, it is especially suited to evaluating how cellular physiology is perturbed by drug candidates. It can be used to study translocation, degradation, secretion, protein–protein interactions and membrane trafficking. The chemiluminescent signal generated does not need dedicated imaging technology and can be read in 96-, 384- or 1,536-well microplates with a standard luminometer.

Tecan's new Infinite series of microplate readers are compatible with its Freedom EVO liquid-handling workstation and are available as either filter- or monochromometer-based systems. According to Feiglin, the sensitivity of the luminescence readers makes them a popular choice for assays of GPCR activity. Using the 1,536-well optics options on the Infinite F500 filter-based microplate reader, a 1,536-well plate can be read in less than 30 seconds. Also new from Tecan is the HydroFlex automated microplate washing and vacuum filtration system, which can be integrated into the Freedom EVO platform.

CyBio of Jena, Germany, made its reputation as one of the first companies to offer multipipettors. It recently unveiled a new robotic liquid-handling platform, CyBi-RoboSpense, which will work with almost any type of sample tube or plate, and can even transport plates to other instruments such as

Tecan's HydroFlex platewasher can be integrated into the Freedom EVO system.

other pipettors, microplate readers or liquid chromatography or mass spectroscopy equipment. The CyBi-RoboSpense can also be used for complicated sample preparation and purification routines such as tryptic digests and magnetic bead separation.

Microfluidics applications

More companies are incorporating the principles of microfluidics into their instruments, making the movement of microlitre — or even nanolitre — volumes of solutions do what used to take hundreds of millilitres, not to mention a saving in time and labour. An innovator in this area is Gyros of Uppsala, Sweden, whose Gyrolab Bioaffy microlabs for protein purification and quantification are based on affinity-capture techniques. The Gyrolab Bioaffy 200 is ideal for high-sensitivity assays to detect low concentrations of proteins. It can be used, for example,

IN SILICO SCREENING WITH CHEMICAL INFORMATICS

Imagine evaluating the efficacy and safety of new drugs without having to lift a pipette. Advances in *in silico* drug development are beginning to make this possible.
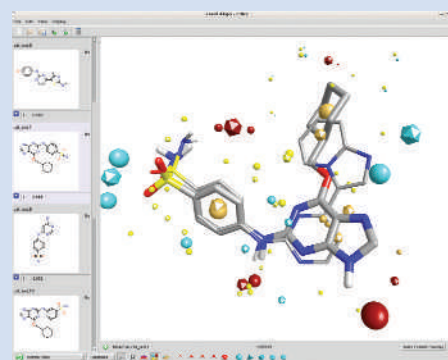
The SYBYL range of computational chemistry software from Tripos of St Louis, Missouri, will design compound libraries, select the best binding partners and screen potential drug molecules for safety and efficacy on the basis of known protein target and/or ligand structures — all *in silico*. A new advanced protein-modelling module predicts the structure of proteins of unknown structure using related proteins of known structure as the models. "It rapidly and accurately provides a valuable protein structure to be used for *in silico* docking experiments," says Kathleen Mensler, vice-president of marketing and corporate development at Tripos. Another new product, Surflex-Sim, helps

find new classes of compounds that are structurally similar to already known, suboptimal, lead compounds, but that may be more active or safer.

Modelling software from Cresset BioMolecular Discovery in Letchworth, UK, uses surface properties or 'molecular fields' of proteins and other molecules to predict intermolecular interactions. "We know that structurally diverse compounds can bind to the same protein site and elicit the same response, but this has never been fully explained by molecular structure," says Sally Rose, Cresset's director of business development. "Our view is that molecular fields provide vital information for understanding and predicting activity," she says. The fields are derived by calculating the interaction energy of a 'probe' atom carrying a positive, negative or neutral charge with the compound. "Our fields summarize the key

binding information, and we locate the most important regions around a ligand where the fields are strongest and binding is expected," says Rose. The original FieldScreen software for virtual screening has been joined by FieldTemplater, designed to find three-dimensional bioactive conformations, and FieldAlign, a molecular-alignment tool for comparing the three-dimensional structures of two molecules.

A pioneer of *in silico* modelling, De Novo Pharmaceuticals in Cambridge, UK, recently announced a new focus for its drug-discovery programme — finding drug targets in various metabolic diseases. De Novo has upgraded SkelGen, its proprietary

Line up: Cresset's FieldAlign program compares molecular structures.

in silico drug-design platform, with the addition of Reflex, which takes into account the rotational flexibility of amino acids in the target's active site. This better evaluates the shape of the active site and its interaction with small-molecule ligands, so that the modelling more closely resembles the dynamics of real protein–ligand interactions.   C.S.

221

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL	
In silico drug design and screening				
Cresset BioMolecular Discovery	FieldScreen for *in silico* screening of lead compounds by their molecular fields, FieldAlign for molecular conformation alignment	Letchworth, UK	www.cresset-bmd.com	
ChemSilico	Software and services for *in silico* ADMETox properties prediction	Tewksbury, Massachusetts	www.chemsilico.com	
De Novo Pharmaceuticals	SkelGen *in silico* drug-design technology	Cambridge, UK	www.denovopharma.com	
IDBS	DAT-LAB and E-workbook electronic lab notebooks, ActivityBase for drug discovery	Guildford, UK	www.idbs.com	
Inpharmatica	Chematica knowledge-based chemogenomics technology platform for target selection through assessment of druggability and identification of hits and leads. Admensa *in silico* and *in vitro* ADMEtox optimization for lead-to-drug-candidate selection	London, UK	www.inpharmatica.com	
Schrödinger	Molecular modelling for *in silico* construction and screening of combinatorial libraries	Portland, Oregon	www.schrodinger.com	
Tripos	Chemical libraries, molecular modelling, and SYBYL screening software for lead candidate discovery	St Louis, Missouri	www.tripos.com	
Automated systems				
Applied Biosystems	DNA sequencers, workstations and automated instruments for genomics and proteomics	Foster City, California	www.appliedbiosystems.com	●
Beckman Coulter	Biomek assay workstations for high-content screening, ADMETox Workstation for *in vitro* membrane permeability measurements	Fullerton, California	www.beckmancoulter.com	
Biacore	Label-free analysis and measurement of biomolecular interactions using surface plasmon resonance	Uppsala, Sweden	www.biacore.com	
Bio-Tek	Automated equipment for liquid handling, microplate automation and microplate reading	Winooski, Vermont	www.biotek.com	
BMG Labtech	Microplate and array readers and handling systems	Offenburg, Germany	www.bmglabtech.com	
Brandel	Automated preparation and harvesting systems, superfusion systems and microdispensers	Gaithersburg, Maryland	www.brandel.com	
Carl Zeiss Jena	Plateexplorer and platescreen systems for ultra-high-throughput screening	Göttingen, Germany	www.zeiss.com	
CyBio	Pipetting, liquid-handling, incubation and imaging systems for automated screening	Jena, Germany	www.cybio-ag.com	
Cytogration	Automated systems for cell culture, membrane growth and *in vitro* drug screening	Rockville, Maryland	www.cytogration.com	
Deerac Fluidics	Equator and Latitude ranges of high-throughput, low-volume liquid-handling systems	Dublin, Ireland	www.deerac.com	
DiscoveRx	PathHunter cell-based assays for intracellular signalling pathways	Fremont, California	www.discoverx.com	
EUGENEX Biotechnologies	Development of test cell lines for drug target screening and cell-based assays	Taegerwilen, Switzerland	www.eugenex.com	
GE Healthcare Bio-Sciences	Instruments, equipment and products for life-sciences research, LEADseeker for lead identification, IN Cell Analyzer confocal imaging system for rapid cellular assays	Little Chalfont, UK	www.gehealthcare.com	
Genetix	Generation, selection and management of mammalian cell lines for drug discovery and biopharmaceutical discovery, colony pickers, arrayers, scanners, and consumables	New Milton, UK	www.genetix.co.uk	
Hamilton Company	Microlab automated liquid-handling workstations	Reno, Nevada	www.hamiltoncomp.com	
Hitachi High-Technologies	Automated instrumentation for mass spectroscopy, liquid chromatography, high-throughput purification, amino-acid analysis and spectroscopy	Tokyo, Japan	www.hitachi-hitec.com/global/index.html	
Innovadyne	Nanodrop automated liquid-dispensing system for low-volume, high-throughput screening	Santa Rosa, California	www.innovadyne.com	
KBiosystems	Robotic instrumentation for high-throughput genomics and proteomics	Basildon, UK	www.kbiosystems.com	
KeyNeurotek	Tissue culture technology and target screening services for drug discovery	Magdeburg, Germany	www.keyneurotek.de	
Labcyte	Pipetting workstations, microplate and magnetic bead washers, DNA extraction	Union City, California	www.labcyte.com	
Magellan Biosciences	ESA Biosciences electrochemical detectors, Dynex microplate readers and automated systems for ADMETox research, TekCel automated sample-storage and handling systems	Chelmsford, Massachusetts	www.magellanbio.com	
OptiCell	Automated cell-culture devices	Westerville, Ohio	www.opticell.com	
PerkinElmer Life Sciences	Cellux cellular fluorescent imaging and Lumilux cell-screening platform for automated high-throughput screening	Boston, Massachusetts	las.perkinelmer.com	
Protedyne	BioCube automated systems for the life sciences and drug discovery, Radius benchtop automated workstations	Windsor, Connecticut	www.protedyne.com	
QIAGEN	BioRobot multifunctional lab workstation and equipment for automated nucleic acid, protein and cell purification	Valencia, California	www.qiagen.com	
Rohasys	Automated equipment for sample pretreatment, weighing, pH, conductivity, turbidity	Rijen, The Netherlands	www.rohasys.com	
RTS Life Sciences	Automated sample management, liquid handling, ultra-high-throughput screening	Manchester, UK	www.rts-group.com	
Scinomix	SciPrep automated sample processor, custom laboratory automation	Earth City, Missouri	www.scinomix.com	
Sirius	Instrumentation and analysis software for pKA, pH and membrane permeability	Forest Row, UK	www.sirius-analytical.com	
Tecan	Freedom EVO automated liquid-handling workstations, Cellerity cell-culture system	Männedorf, Switzerland	www.tecan.com	●
The Automation Partnership	Automation for liquid handling, compound storage, high-throughput screening	Royston, UK	www.automationpartnership.com	
Thermo Fisher Scientific	ArrayScan HCS and KineticScan HCS systems for automated cell-based screening of candidate drugs	Pittsburgh, Pennsylvania	www.cellomics.com	●

COMPANY	PRODUCTS/ACTIVITY	LOCATION	URL	
Tomtec	Liquid-handling workstations	Hamden, Connecticut	www.tomtec.com	
Torcon Instruments	Precision fluid-dispensing and automated microplate-handling equipment	Torrance, California	www.torconinstruments.com	
Universal Technology	Automated storage and retrieval systems for the pharmaceutical industry	Pittsburgh, Pennsylvania	www.univtech.com	
Velocity11	Automated high-throughput liquid-handling, microplate management, dispensers and pipetting stations	Palo Alto, California	www.velocity11.com	
Warner Instruments	Electrophysiology and cell-biology equipment and products	Hamden, Connecticut	www.warneronline.com	
Wave Biotech	Equipment for cell culture, Wave Bioreactor disposable cell-culture systems	Bridgewater, New Jersey	www.wavebiotech.com	
Xiril	Liquid-handling platforms and homogenization equipment	Hombrechtikon, Switzerland	www.xiril.com	
Zinsser Analytic	Pipetting, extraction and liquid-handling robots	Frankfurt, Germany	www.zinsser-analytic.com	

Microfluidics

Caliper Life Sciences	LabChip microfluidic systems for high-throughput screening	Hopkinton, Massachusetts	www.calipertech.com	
Fluidigm	Topaz microfluidics chips for protein crystallization, BioMark microfluidics systems for qPCR and detection of rare sequences	South San Francisco, California	www.fluidigm.com	
Gyros	CD microlaboratories for protein quantification at the nanolitre scale	Uppsala, Sweden	www.gyros.com	

Ion-channel screening

Aurora Biomed	Ion Channel Reader for voltage- and ligand-gated channels, assay services, cell lines, automated liquid-handling systems	Vancouver, British Columbia	www.aurorabiomed.com	
Aviva Biosciences	Equipment for electrophysiology, SEAL Microchips for high-throughput patch-clamp ion-channel measurements	San Diego, California	www.avivabio.com	
Cellectricon	Patch-clamp equipment for ion-channel screening	Gothenburg, Sweden	www.cellectricon.se	
ChanTest	Detection of intracellular ion-channel trafficking	Cleveland, Ohio	www.chantest.com	
Cytomyx	Cell lines and cDNAs for kinases, receptors and ion channels	Cambridge, UK	www.cytomyx.com	
flyion	Flyscreen high-throughput patch-clamp system	Tübingen, Germany	www.flyion.com	
Invitrogen	Voltage sensor reagents, kits and reagents for genomics, proteomics and cell biology	Carlsbad, California	www.invitrogen.com	●
Molecular Devices	PatchXpress and IonWorks for automated patch clamping, substrate finder for protein kinases, automated liquid-handling systems	Sunnyvale, California	www.moleculardevices.com	●
Multi Channel Systems	Automated equipment for electrophysiology, *Xenopus* oocyte injection, ion-channel screening	Reutlingen, Germany	www.multichannelsystems.com	
Nanion Technologies	Port-a-patch benchtop patch-clamping system for voltage- and ligand-gated ion channels	Munich, Germany	www.nanion.de	
Sophion Bioscience	QPatch 16 for automated patch clamping	Ballerup, Denmark	www.sophion.dk	

General

Agilent	Instrumentation, kits and consumables for the life sciences	Palo Alto, California	www.agilent.com	
Bio-Rad	Products, instruments and software for life-sciences research	Hercules, California	www.bio-rad.com	●
Cambrex	Gels and other products for molecular and cell-biology research	East Rutherford, New Jersey	www.cambrex.com	
Charles Rivers Laboratories	Services for biopharmaceutical and preclinical research	Wilmington, Massachusetts	www.criver.com	
Eppendorf	Laboratory instrumentation and consumables for molecular and cell biology	Hamburg, Germany	www.eppendorf.com	●
Gilson	Liquid-handling and pipetting instruments	Middleton, Wisconsin	www.gilson.com	●
Millipore	Automated equipment for molecular biology, biochemistry, genomics and proteomics	Bedford, Massachusetts	www.millipore.com	
Nalge Nunc International	Labware	Rochester, New York	www.nalgenunc.com	
Pierce Biotechnology	Components and consumables for cell and molecular biology	Rockford, Illinois	www.piercenet.com	●
Pro Scientific	Laboratory equipment, centrifuges	Oxford, Connecticut	www.proscientific.com	
R&D Systems	Reagents and assays for life-sciences research	Minneapolis, Minnesota	www.RnDSystems.com	●
Roche Applied Science	Reagents and kits for molecular biology, functional genomics and proteomics research	Indianapolis, Indiana	www.roche-applied-science.com	●
Sigma-Aldrich	Reagents and biochemicals for life-sciences research	St Louis, Missouri	www.sigmaaldrich.com	●
Stratagene	Tools and reagents for molecular biology, genomics, proteomics, drug discovery	La Jolla, California	www.stratagene.com	
TaKaRa Bio	Reagents, kits and custom services for life-sciences research	Shiga, Japan	www.takara.bio.com	●
Upstate	Antibody production, reagents for signalling research	Charlottesville, Virginia	www.upstate.com	
Wako Chemicals	Speciality chemicals, bioproducts and clinical diagnostic reagents	Richmond, Virginia	www.wakousa.com	

● see advertisement

naturejobs

Scientists often describe themselves as a community. If this description is true, then that community has grown "nastier and nastier" in the past 30 years, according to one participant at a retreat for US National Cancer Institute postdocs last week in Ocean City, Maryland. The remark came during a workshop on scientific publishing that I took part in. The nastiness, she said, manifests itself most often in the form of hostile, disparaging remarks in referee reports during the peer-review process — usually cloaked in anonymity.

The culture of competition for limited resources has probably contributed to this problem, said my co-panellist Feng Chen, senior editor at *Molecular Cell*. But she noted that rejected authors can also fall into this trap — sending angry e-mails or making emotional phone calls just minutes after their manuscripts have been sent back. She recommends that would-be authors take a 24-hour 'cooling-off' period before they respond to journal editors if their paper is rejected. Authors can also shield themselves from some hostility by requesting that colleagues who bear them a grudge be excluded from the review process — although requesting that reviewers from an entire continent or discipline be barred may not be entirely reasonable, Chen added.

Perhaps scientists should remember that they are indeed part of a community, and that this comes with an unwritten social contract, to be at the very least cordial and professional towards each other — even when their names are hidden. Editors could encourage reviewers to provide constructive feedback, even when they don't recommend acceptance of a paper. Telling the would-be author what works in the manuscript and what still needs to be done is more beneficial than making anonymous insults.

Perhaps both reviewers and editors should have a 24-hour 'cooling off' period, when they can think about how they would communicate their intended message to a neighbour or a friend. *Nature*'s advice and policies for authors and reviewers on these and other related matters can be found at www.nature.com/authors.

Paul Smaglik, *Naturejobs* editor

Breaking with tradition

There's more than one way to do postdoctoral research, and unconventional routes can lead to international collaboration, intellectual or personal freedom and better job prospects. Kendall Powell explores the roads less travelled.

Not long after Justin Brashares arrived in the United Kingdom for a postdoctoral fellowship in wildlife ecology at the University of Cambridge, he found himself in unexpected cross-cultural crossfire. As he quietly enjoyed a sandwich at his computer, two colleagues walked by his office door and did a double-take.

"They were staring, horrified, like I had dragged a raw sheep carcass into my office and was eating it at my desk," recalls Brashares, who was sponsored by the US National Science Foundation (NSF) International Research Fellow Program. His new labmates eventually convinced him to take a break for lunch — and for tea twice a day. Although moving from the United States to Britain for a postdoc had seemed like a small shift, Brashares soon learned the benefits of working in a different scientific culture.

The time-tested track for academic postdocs, remaining in their home country to work within one research supervisor's lab group, is not the only one. In some fields or for some individuals, traditional opportunities prove limiting. Seeing science done differently is just one of the advantages of following an alternative route. Some idiosyncratic opportunities can open doors to international mobility, personal flexibility and novel research projects. Venturing a little farther afield brings challenges too — but it can propel you into enviable permanent positions.

Globetrotting

Now an assistant professor of animal ecology at the University of California, Berkeley, Brashares split his time as a postdoc between fieldwork with the Wildlife Division of the Forestry Commission of Ghana and data analysis at Cambridge. The international contacts and independence he gained made him a stronger job candidate — he received two job offers in his first fellowship year. Recruiting departments, he says, "want to see that you are not just churning out things that you peeled off your adviser's vision".

International postdocs have their hurdles, though,

Galen Halverson, top, and Andrea Decker: moving on.

particularly in adjusting to a different rhythm of work. A couple of years ago in Ghana, Brashares had only dial-up access to the Internet — and that was just once a week. "You can either explode or learn to deal with it," he says — and appreciate the break from e-mail.

Galen Halverson, another former NSF international research fellow, adds that switching countries means necessary downtime after the relocation. He moved from Boston, Massachusetts, to Toulouse, France, for a postdoc at the Laboratory for Transfer Mechanisms in Geology (LMTG). Not only did he have to overcome language barriers and adjust to a much more laid-back pace than he was used to, but it took about eight months before he was analysing data again.

"Allot yourself a good chunk of time to get going," says Halverson, now a lecturer in geology at the University of Adelaide, Australia. Consider doing a slightly longer postdoc, he advises, or bring along some unfinished work to make best use of that downtime.

International fellowships can also be a route to gaining specialized technical expertise. At the LMTG, Halverson learnt how to use a relatively new kind of mass spectrometer for measuring isotopes. That knowledge was key to his current department colleagues, who want to acquire a similar instrument.

Charles Tahan took his postdoctoral collaborations in quantum devices to an extreme, working at the University of Melbourne, Australia, the University of Tokyo, Japan, and the University of Cambridge, UK. Through this continent-hopping, Tahan gained in two years contacts that would normally take a decade.

But with just a limited time at each place, he could not have followed Halverson's advice. A peripatetic postdoc must be "physically and mentally ready to start a project right away" in order to actually finish anything, he says. If you want to take on something so ambitious, he recommends being extremely well read in the field you are entering before you show up.

There are definite downsides to extreme mobility, too. Tahan says he had to rely on his parents in the United States to take care of important mail and

logistics such as figuring out taxes, healthcare and banking as he bounced around. He also concedes that he would probably have published more papers if he had stayed in one place. But he feels the trade-off was worth it: "I gathered more new ideas in my back pocket that I can build on in the future."

In the European Union (EU), Marie Curie fellowships allow postdocs to move within Europe or between the EU and other countries. They come with a bonus: the Marie Curie Fellows Association (MCFA), which has more than 3,000 current and former fellows as members and so forms a powerful cross-disciplinary and multicultural network. The MCFA runs projects to improve the working conditions of all young researchers in Europe and can help sort out common problems that arise for mobile fellows, says Vanessa Diaz, vice-chair of the MCFA and a postdoc at the University of Sheffield, UK.

"Go and see the grass on the other side of the fence," encourages Diaz, an Italian who did her first degree in Venezuela and her doctorate in France. It shows you "the extraordinary richness of human society, which will hopefully translate into more exciting research".

Several international postdocs say their advisers or labmates warned against leaving their home country's system if they wanted to return to a permanent position there. But none regrets the decision or struggled to find a job later. On the contrary, most say their fellowship opened more doors than it closed.

Go mentorless

The doors are continually revolving at two unique centres for synthesis research: the National Center for Ecological Analysis and Synthesis (NCEAS), part of the University of California, Santa Barbara, and the National Evolutionary Synthesis Center (NESCent) in Durham, North Carolina (see Table, overleaf). Fellows rub elbows with hundreds of scientists each year.

"The opportunity to meet a lot of different people is really powerful," says Melinda Smith, former NCEAS postdoc and now a plant-community ecologist at Yale University. "It exposed you to how science is done successfully in many ways." That's a big advantage for job searching, she says.

The flipside is that these positions are mentorless. Fellows have no lab group or research adviser. Such independent research is a boon when it's time to move on: 84% of NCEAS postdocs go on to tenure-track or equivalent positions. But it can be daunting early on.

"It is hard to get over the hump and get started," says Kirsten Fisher, a current NESCent fellow. "It helps to have a monthly game plan." Fellows who are self-motivated will flourish, although some find it useful to set deadlines with collaborators.

Both centres have extra resources for computing and statistical help. Postdocs learn from each other: Fisher says she could have picked up bioinformatic techniques more quickly had she asked colleagues sooner. To prevent any feeling of working in a vacuum, NESCent encourages fellows to associate with a lab group from one of three local universities, and the NCEAS provides an allowance to visit faraway mentors.

These centres are not for everyone. They emphasize synthesis research and discourage gathering new

Elsa Cleland: gathering skills that will last for a whole career.

"You need to cross-pollinate with as many researchers as possible. It will give you a more complete view of the big questions."
— Justin Brashares

primary data. Fellows use raw data gathered by others, or crunch numbers from published studies in novel ways. This would not appeal to anyone craving fieldwork or wishing to learn about a new organism, ecosystem or technique.

But they promote career-development skills such as project management and collaboration. Many of the postdocs are operating like academic faculty members — and recruiters seek people who can do this. "Every day, I consult someone and ask their opinion about something I'm doing," says postdoc Elsa Cleland. "That's what you do for the rest of your career."

These researchers may lack the camaraderie of a lab group, but many say they feel less isolated, with their cohort of fellows, than in a traditional postdoc position.

"It's a different niche that wasn't available before," says Kathleen Smith, director of NESCent. "And for a lot of postdocs it's perfect."

Beyond the lab

Another set of fellowships may suit young scientists who want to teach as well as research, or who are looking for flexible work schedules or a bridge between academia and industry. They are harder to find than standard postdoc positions, but could help you find your own perfect niche in the research enterprise.

SPIRE fellows at the University of North Carolina at Chapel Hill spend two years in the lab and one year teaching undergraduate students at other universities in the state that have historically drawn most of their students from minority ethnic groups. The National Institute of General Medical Sciences also has such programmes in California, Arizona, Georgia, Kansas and Tennessee.

Designed for those committed to both teaching and research, SPIRE (Seeding Postdoctoral Innovators in Research and Education) has an amazing track record: every one of its graduates has gone on to the position of their choice, with 70% in a tenure-track position.

Katie Shannon, a former participant and now a cell biologist at the University of Missouri-Rolla, was looking for a programme that valued teaching. Through SPIRE she got hands-on lessons in pedagogy and built a portfolio that placed her on several job shortlists.

Because the research portion is limited to two years, Shannon advises choosing a relatively well developed project and an adviser who is supportive of your decision to teach. She cut her preparation time for giving a lecture from six hours to two during the fellowship, which gave her a head start in Missouri.

"There is a big market out there for fellows who want to stay in science and are desperately looking to get the professional teaching skills they also need," says programme director Leslie Lerea.

Another fellowship targets postdocs who might otherwise leave science by giving them more flexible work schedules. The Dorothy Hodgkin Fellowships, sponsored by the Royal Society in the United Kingdom, support young scientists who need to care for children or other relatives, or who have a medical disability.

The fellowship guarantees four years of funding that can be used for up to six years of research. In other words, fellows are free to take a leave of absence or work part-time for a spell and make up that time at the

end of the fellowship.

"It's not always possible to be productive on a weekly or monthly basis," says Deirdre Black, a theoretical particle physicist at the University of Cambridge and mother of a young son. Having four or five years in this scheme to build up a body of research will keep her competitive for permanent positions with people who haven't been rearing young children, she says.

Fellowship	Number of annual places	Field of research	Duration	Annual salary	Other funds and resources
NCEAS	12–18	Ecology	2 to 3 years	$42,000 to $44,000	$2,500/year research funds, $2,000 mentor allowance, access to Berkeley resources
NESCent	7–8	Evolutionary biology	2 to 3 years	$37,000 plus benefits	$5,000 research funds, connections to local universities
SPIRE	4–7	Any area funded by NIGMS	3 years	$37,000 and up (NIH payscale)	Allowances for research, travel, computer
Dorothy Hodgkin	About 10	Any area funded by the Royal Society	4 to 6 years	Host institution's payscale	£11,000–13,000/year research funds; childcare allowance during travel
NSF International Research	30–35	Any area funded by the NSF	9 months to 2 years	$30,000 to $60,000	Allowances for relocation, health insurance, professional travel, and research funds

After her maternity leave, Black returned to work at 40% for a few months and now works 80%, or slightly shorter days, five days a week. That way, she can participate in her research group every day and still have more time with her son.

For those wondering which side of the academia–industry fence is greener, there are opportunities to hop over without committing yourself. Most postdocs who have worked in industry say you can almost always go back to academia if you want to badly enough. It's the person who never ventures into the corporate world at all who misses out on exploring a range of job options.

But there are some industry postdocs that definitely make the crossing easier. Take the Marie Curie 'transfer of knowledge' fellowships, which allow for the exchange of postdoctoral researchers between academic and industrial institutions in the EU.

Antonella Di Trapani did her postdoctoral work in nanomaterials with chemicals company Johnson Matthey in Sonning Common, UK. She learned skills in sales, marketing and applied science, and liked the tightly focused research. Changing projects every few months appealed to her, as did having the resources to do more efficient experiments.

"The world seems so small in academia, but once you are in industry, the opportunities open up," she says. She followed her interest in research management to become programme coordinator for materials science at the

Kathleen Smith: some want a new and different niche.

European Science Foundation in Strasbourg, France.

Andrea Decker still cannot decide between tenure-track or high-tech, but luckily her postdoc position at IBM's Zurich Research Laboratory in Switzerland leaves all options open. Decker does early-stage exploratory research on nanotechnology, publishes her findings and presents her work at conferences — in contrast to most industry postdoc fellows.

Still, she acknowledges the danger of losing contact with the academic world. She encourages industry postdocs to seek work on the large collaborative projects set up in the European Research Area framework. Meetings provide ample opportunity to network with academic researchers from across the EU.

"Move about and get to know different cultures," Decker advises — both scientifically and globally. Brashares agrees; he stays in daily contact with UK and West African collaborators. Busy setting up his own group at Berkeley, he is back to eating his lunch over the keyboard. But he hasn't forgotten the lessons his unconventional postdoc taught him. "People applying for these positions should be gluttons for adventure," Brashares says. "They want to move around and cross-pollinate with as many researchers as possible. It will give them a more complete view of the big questions and different approaches to answering them." ■

Kendall Powell is a freelance science writer based in Broomfield, Colorado.

Web links

Australia Research Council
❧ **www.arc.gov.au**
Dorothy Hodgkin Fellowships
❧ **www.royalsoc.ac.uk/funding.asp?id=1122**
European Molecular Biology Organization
❧ **www.embo.org/fellowships/index.html**
Japan Society for the Promotion of Science
❧ **www.jsps.go.jp/english/e-fellow/fellow.html**
IBM Zurich Research Laboratory
❧ **www.zurich.ibm.com**
Marie Curie Fellows Association
❧ **www.mariecurie.org**
Marie Curie Actions
❧ **cordis.europa.eu/mariecurie-actions**
National Center for Ecological Analysis and Synthesis
❧ **www.nceas.ucsb.edu**
National Evolutionary Synthesis Center
❧ **www.nescent.org**
NSF International Research Fellowship Program
❧ **www.nsf.gov/funding/pgm_summ.jsp?pims_id=5179**
SPIRE
❧ **spire.unc.edu/links/iracda.html**

C. BURTON

Justin Brashares: ideal work if you're a glutton for adventure.

The inside track from academia and industry

Follow the money

Borders no longer constrain the flow of capital, ideas and talent to emerging centres of excellence.

Michael Alvarez

Global economics is changing and science is changing along with it — a trend that has implications for everyone from students to industry researchers to policy-makers in federal and local government. The convergence of communications and computing advances has brought new players into the global marketplace. As a result, competition for both talent and market share is intensifying, and the field has been levelled.

This levelling effect poses a challenge for traditional science and technology companies and the regions that host them. They need innovators and talented employees in order to stay competitive.

It also creates job opportunities in the emerging centres. As the field of players grows more crowded and diverse, there are greater opportunities for new countries, institutions and individuals to emerge as leading producers of cutting-edge innovations. A growing middle class in developing countries, including well-educated and technically skilled people, means that there is a greater base of individuals both generating and circulating money for their local economies.

The competitive advantage belongs to those companies, regions and countries most capable of transforming technologies into commercial products. These innovations provide them with both trade opportunities and sustainable employment. The transformation requires capital, infrastructure, ground-breaking technology — and a high-quality workforce.

The quest for technologies and talents is less bounded by geography than in the past. The capital, technology and skilled professionals needed to drive

> **"Regions seeking to become 'the' hub of any innovation need talented individuals to help them attract investment."**

economic growth today can be gathered from numerous locations, and companies are increasingly able to attract the resources they need from distant corners of the globe. Note, for example, that the four leading institutions in terms of biotech publication rankings are located in three different countries (see 'Turning ideas into income').

Competition is keen among regions, both within countries and internationally, to become known as 'the' international hub for selected high-technology sectors. It's not just India and China — European cities and non-traditional research locales in the United States, from Ohio to Florida, are competing for talent and funds. They position themselves as being hospitable to business, by providing inducements such as capital, access to markets, favourable regulatory environments and, of course, a highly skilled technical workforce.

One prime example is embryonic stem-cell research. The United States' restrictive federal policy has held it back while countries such as Australia, the United Kingdom and Singapore (and some US states, mostly notably California) aim to become hubs for a field that is still in its infancy. Across many areas of science and technology, new players have emerged to compete with traditional centres. From Bangalore to Berlin and Shanghai to San Diego, more and more cities are aiming to become leaders in promising high-technology sectors.

So how can one identify regions that actually have what it takes to succeed in this global marketplace, where the opportunities for sustainable profits, employment and technological development are greatest?

Although computing and telecommunications technologies have played down the importance of geographical location, it is crucial for companies to be located close to universities and to collaborate with them.

Universities play a leading role in generating early-stage technologies and producing skilled trainees. They also patent and license technologies for commercial development. This means that investors and researchers from across the globe can be on the spot to engage with the technological advancements and people most relevant to their own objectives and interests.

The onus is also on industry and government to provide policies, infrastructure and funds to translate ideas and discoveries into products. Their moves in this regard may very well determine the success or failure of science and technology hubs.

The regions that will earn a place at the front of the pack are those most capable of adapting to the new playing field, operating effectively within a distributed work environment and developing relationships across borders to source the best talents and technologies. Skilled individuals who are flexible and mobile enough to take advantage of the regions' needs have a good chance of reaping the rewards. ■

Michael Alvarez is director of Stanford University's School of Medicine Career Center, California.

TURNING IDEAS INTO INCOME

Biotech publication rankings
1. Harvard University
2. University of Tokyo
3. University of London
4. University of California, San Francisco

Biotech patent ranking
1. University of Texas
2. University of California, San Francisco

3. Johns Hopkins University
4. Stanford University

Technology-transfer and commercialization index
1. Massachusetts Institute of Technology
2. University of California
3. California Institute of Technology
4. Stanford University

Source: Milken Institute, 2000–04

Discover True Innovation

Director, Academic Scientific Liaison

Merck Sharp & Dohme Limited (MSD) is the UK subsidiary of Merck & Co., Inc., of Whitehouse Station, New Jersey, USA, a leading research-based pharmaceutical company that discovers, develops, manufactures and markets a wide range of innovative pharmaceutical products to improve human health.

In a world where drug discovery is no longer the exclusive domain of large pharmaceutical companies, MSD understands that both the needs of patients and its own future are served best by teaming up with others. That's why in recent years MSD has rapidly increased its investments in partnership deals with biotech start-ups, venture capital firms and academic researchers to find promising technologies, drug targets and candidate products in their very early stages.

Licensing & External Research, Europe, based at Hoddeson UK is currently looking for a Director, Academic Scientific Liaison to work across the European region as well as Israel and South Africa. The holder of this new position will work in concert with our existing four regional directors and with colleagues in all MSD / Merck divisions to ensure that we are aware of technology and drug discovery collaboration opportunities arising in the wider academic community. The success of this position will result in successful entry into exciting breakthrough fields.

This high profile and stimulating role will be responsible for building relationships leading to mutually beneficial basic research collaborations and strategic alliances with academic groups working at the cutting edge of science relevant to MSD / Merck's priority research areas. Reporting to the Executive Director of Licensing and External Research, Europe, the successful candidate will possess a PhD in a bioscience discipline and / or an MD degree and will have extensive knowledge of basic research gained through laboratory experience and publication. A thorough knowledge of basic research technologies as applied to medical / drug discovery / biomarker targets and a broad-based scientific understanding across human health therapeutic areas are also essential.

In order to liaise effectively with target contacts, the role requires exceptional inter-personal and excellent communication skills with the ability to lecture confidently to audiences at major international conferences. A positive and flexible attitude with good attention to detail and a willingness to learn new skills is essential. Fluency in English is essential and fluency in a second major European language would be advantageous. The position will require frequent travel mostly within Europe. We are willing to consider candidates from both industrial and academic backgrounds.

In return for your commitment and enthusiasm we will offer an attractive and competitive benefits package including a company car.

If you are interested in applying for this position, please send your Curriculum Vitae with a covering letter, quoting reference: Licensing, to the Human Resources Department, MSD UK, Hertford Road, Hoddesdon, Herts EN11 9BU. Alternatively you can email your application to hoddesdonrecruit@merck.com

For more information, please visit our websites: www.msd-uk.co.uk and www.merck.com

MERCK SHARP & DOHME

98170RM
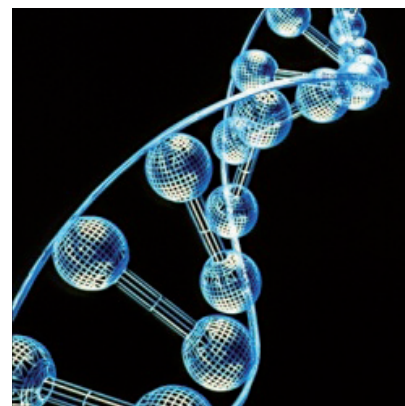
DIRECTOR

**International
Computer Science Institute
Berkeley, CA**

The International Computer Science Institute (ICSI), an independent non-profit laboratory closely affiliated with the Electrical Engineering and Computer Science Department at the University of California, Berkeley (UCB), invites applications for the position of Director, to begin in the Fall Semester of 2008.

The ICSI Director's primary responsibilities are to: oversee and expand ICSI's research agenda; act as a high-level external evangelist for ICSI research; identify and pursue strategic funding opportunities; and strengthen ICSI's relationship with UCB. The Director reports directly to ICSI's Board of Trustees.

ICSI is recognized for its world-class research activities in speech and language processing, networking, and computational biology, and is building a new competency in computer architecture. Several of ICSI's research staff have joint appointments at UCB, and many UCB graduate students perform their research at ICSI. In addition, ICSI places significant emphasis on international partnerships and visiting scholar programs with countries in Europe and Asia.

ICSI is seeking a Director with sufficient breadth, interest and professional connections to promote and augment ICSI's ongoing research efforts. Applicants should have recognized research leadership, as well as a strong record in research management and demonstrated success at government and industrial fundraising. Experience with international collaboration and fundraising is a plus.

Applications should include a résumé, selected publications, and the names of three references who will send recommendations. Review begins May 1, 2007; candidates are urged to apply by that date.

To learn more about ICSI, go to **http://www.icsi.berkeley.edu.**

To apply for the Director position, send the above material to
apply@icsi.berkeley.edu

Recommenders providing letters should send them directly to apply@icsi.berkeley.edu by 7/1/2007. ICSI is an Affirmative Action/Equal Opportunity Employer. Applications from women and minorities are especially encouraged.

NW97192R

THE UNIVERSITY OF FLORIDA
COLLEGE OF MEDICINE

Chair, Department of Neuroscience

The University of Florida College of Medicine invites applications and nominations for the position of Professor and Chair in the Department of Neuroscience. The department has a mission within the college of both research and teaching responsibilities. The department has 24 primary faculty members with research interests in all aspects of neuroscience (**www.neuroscience.ufl.edu**). The department is housed within the McKnight Brain Institute (**www.mbi.ufl.edu**), a state-of-the-art facility for research and teaching in neuroscience and neuroscience-related disciplines. The successful candidate will have a Ph.D., M.D., or M.D./Ph.D. degree. We seek an accomplished scholar with a distinguished record of research, international recognition, leadership, and administrative skills. The review of applications will begin on March 30, 2007 and will continue until the position is filled. Applicants should provide a letter of application, curriculum vitae and the names of three references via email to **NS-search@phys.med.ufl.edu** or by mail to: **Charles E. Wood, Ph.D., Chair, Search Committee, Box 100274, Department of Physiology and Functional Genomics, University of Florida, Gainesville, FL 32610-0274.**

The University of Florida is an Equal Opportunity Institution.

NW97558R

President and Director
Woods Hole Oceanographic Institution
Woods Hole, Massachusetts

Woods Hole Oceanographic Institution (WHOI), the world's largest private institution dedicated to research and education at the frontiers of oceanography, seeks a dynamic new President and Director. The Institution supports world-class research and education with vessels and instruments that enable unmatched access to the sea and with premier shore-based laboratories and other facilities. With an annual budget of $136 million, the WHOI community includes some 500 scientific and technical staff, nearly 400 operating, administrative, and support staff, and approximately 200 students and postdoctoral scholars.

This is a rare opportunity to lead an institution with a remarkable history of achievement and the resources to continue to expand understanding of the oceans and their central influence on Earth systems and human society. The new President and Director will be a person of exceptional intellectual vision and scientific judgment, a proven leader who listens well and communicates persuasively, and a person who embraces the pressing scientific work of the Institution with passion.

WHOI is working with a national executive search firm, Isaacson, Miller, on this recruitment. For a detailed position announcement or to apply, please email 3350@imsearch.com. All inquiries will be held in strict confidence.

WHOI is an Affirmative Action/ Equal Opportunity Employer.

1930

NW97615R

VCU
Virginia Commonwealth University

TENURE-TRACK FACULTY POSITIONS
Chemical and Life Science Engineering

The Department of Chemical and Life Science Engineering (CLSE) in the School of Engineering at Virginia Commonwealth University (VCU) has two tenure-track faculty openings starting in the Fall of 2007. One position is at the **Associate level**, and the other is at the **Assistant level**. Chemical and Life Science Engineering at VCU represents the broad, formal interaction of the disciplines of chemical engineering with life and health sciences to create a forward-looking, nationally distinct program. Many of the Life Science areas at VCU enjoy national rankings, including those in the medical sciences, biological sciences and environmental life sciences. The new School of Engineering formed in 1996 has embarked on a "25 in 25" initiative to become a top 25 program in 25 years. Notable facilities include the new 120,000 sq ft School of Engineering building, the new Trani Center for Life Sciences, the Rice Institute for Environmental Life Sciences located along the coastal plain region of the James River, and VCU's Medical School and Hospitals. Current research areas in CLSE include stem cell and stem-cell-derived tissue engineering, cellular engineering and signal pathway analysis, biological systems engineering, bioinformatics and biocomputing, genetic and protein molecular engineering, small molecule and cellular based therapeutics, reaction engineering and molecular transport, advanced polymeric materials and processing methods. A new research Institute for Health and Life Science Engineering and a new Phase II Engineering building expansion will be open in 2007-2008 that will greatly expand and enhance research and education capabilities in the life science engineering areas. Candidates must have earned a Ph.D. and at least one degree in Chemical Engineering or Bioengineering or closely related discipline.

Outstanding candidates should submit a complete curriculum vitae, statement of research and teaching interests, and a list of four references to: **Dr. Michael H. Peters, Chair, Chemical and Life Science Engineering, Virginia Commonwealth University, 601 West Main St., Room 403A, P.O. Box 843028, Richmond, VA 23284-3028**. Electronic submissions are acceptable by *.pdf files only please to: **jbschrei@vcu.edu.** Candidates must be eligible for employment in the United States by indicating their citizenship or visa status. Review of applications will continue until the positions are filled.

*VCU is an Equal Opportunity, Affirmative Action Employer.
Women, minorities and persons with disabilities are strongly encouraged to apply.*

NW97818R

UNIVERSITY OF HAWAI'I
DEAN, COLLEGE OF NATURAL SCIENCES

The University of Hawai'i at Mānoa (UHM), a leading institution of higher learning in the Pacific Basin and one of the nation's few land-, sea- and space-grant institutions, seeks an innovative, entrepreneurial and highly accomplished leader as Dean of the College of Natural Sciences.

The College of Natural Sciences is part of an international community of scholars in the physical and life sciences, mathematics and information and computer sciences, including library sciences. The Dean serves as the foremost academic and administrative leader for the College. As the University celebrates its centennial in 2007, the Dean will be called upon to develop an inspiring cohesive vision and strategic plan to guide the College toward new levels of excellence by building on the strong scientific accomplishments of existing programs and by pursuing an integrated, comprehensive academic program that balances scholarship, instruction and service. To implement this vision, the Dean must be a passionate, collaborative, engaging advocate for the College within the University, across the state, and beyond.

Nominations and applications are being accepted for the position. First screening of candidates will begin on Monday, March 19, 2007, and will continue until the position is filled. Candidates must submit a cover letter summarizing the candidate's interest and qualifications for the position, a current résumé, and the names of five (5) professional references including postal and e-mail addresses and telephone numbers. For more information about the University of Hawai'i at Mānoa, please go to www.uhm.hawaii.edu For a position description, application and nomination requirements, and other information, please visit University of Hawai'i web sites at http://workatuh.hawaii.edu and www.hawaii.edu/executivesearch/naturalsciences

Inquiries, nominations, and applications should be directed to Edward W. Kelley and Partners at:

Edward W. Kelley and Partners, Attention: Sharon Tanabe, Principal
1111 Corporate Center Drive, Suite 106
Monterey Park, CA 91754
Phone: 323-260-5045
Fax: 323-260-7889
E-Mail: sharon.tanabe@ewkp.com

E-Mail correspondence is strongly encouraged

The University of Hawai'i is an equal opportunity/affirmative action institution and encourages applications from and nominations of women and minority candidates.

NW97961R

JOHNS HOPKINS
MEDICINE

DIRECTOR

DEPARTMENT OF NEUROLOGY
JOHNS HOPKINS SCHOOL OF MEDICINE
NEUROLOGIST-IN-CHIEF, THE JOHNS
HOPKINS HOSPITAL

The Johns Hopkins University School of Medicine and The Johns Hopkins Hospital are seeking an exceptional individual to lead the Department of Neurology who is dedicated to discovery and innovation in clinical care, research and teaching and who demonstrates outstanding leadership abilities. The successful applicant must share our unwavering commitment to excellence, integrity, collegiality and respect for and inclusion of individuals of diverse backgrounds. The Director will play a critical role in reshaping the Neuroscience landscape at Johns Hopkins including collaboration in the formation of a new Brain Science Institute and biotech park.

Please send letter of application, curriculum vitae and bibliography to:
John A. Ulatowski, MD, PhD, MBA
Chair, Neurology Search Committee
Johns Hopkins University School of Medicine
733 North Broadway, SOM #100 Baltimore, MD 21205 Or email your information to:
neurologysearch@jhmi.edu

An Affirmative Action/Equal Opportunity Employer

NW98145R

Corrosive, uncompromising opinions and views on workplace issues affecting YOU.

naturejobs

Director
Dell Pediatric Research Institute
The University of Texas at Austin

The University of Texas invites applications and nominations for the position of Director of the Dell Pediatric Research Institute (DPRI). The successful candidate will play a pivotal role in creating this major new biomedical research institute. DPRI will be the first of several institutes planned for a new Health Research Campus at UT Austin located approximately three miles from the main UT Austin campus near downtown Austin within a 711 acre, new urban village. The DPRI will be housed in a new 150,000 ft2 building with capacity for 28 senior faculty and their research programs. The DPRI Director will be expected to play a major role in recruiting these faculty; generous compensation and start-up packages will be available. We expect strong interactions between DPRI scientists and those on the UT Austin main campus.

We are seeking a dynamic leader with demonstrated experience in administration and communication, ability to lead a major biomedical research unit, and ability to interact well with both private supporters and public education leaders in the state. Additionally, he or she must have an earned Ph.D. or M.D. with a demonstrated record of accomplishment in biomedical research. The Director will report to the Executive Vice President and Provost and the Executive Vice Chancellor for Health Affairs through the Dean of the College of Natural Sciences.

The University of Texas at Austin is the flagship university of the UT System's 15 academic and health institutions. Among the top research universities in the US, UT Austin is home to almost 50,000 students, 2,700 faculty and 17,000 staff members, the largest graduate program in the nation, and one of the largest total student enrollments. The University received more than $300 million in grants and contracts, had more than 96 patents awarded last year, and has recently completed a successful seven-year capital campaign resulting in gifts totaling nearly $1.55 billion.

Please submit a letter of nomination or interest and for the latter, a statement of experience and curriculum vitae to **Dr. Mary Ann Rankin, College of Natural Sciences G2500, 1 University Station, The University of Texas, Austin, Texas 78712.**

For additional information please see **http://cns.utexas.edu/dpri**

The University of Texas is an Equal Opportunity Employer. Qualified women and minorities are encouraged to apply.

NW97198R

Department for Environment, Food and Rural Affairs

Chief Scientific Adviser

£150K pro rata

The Department for Environment, Food and Rural Affairs (Defra) wishes to appoint a Chief Scientific Adviser in succession to Professor Sir Howard Dalton FRS who will be stepping down during the Summer of 2007 at the end of his term of office.

As the lead government department for a number of vitally important policy areas ranging from climate change, through ecosystems, to avian influenza, Defra has a duty to ensure that its policies are based on robust scientific evidence. A recent Office of Science and Innovation Review has placed Defra as a market leader in using science in policy making as well as identifying a number of areas where it demonstrates best practice. These include its development of an overarching Evidence and Innovation Strategy, its success in horizon scanning and in improving the quality and relevance of the science it commissions; and in its open and transparent working practices.

The Chief Scientific Adviser is one of the most senior and influential posts within Defra playing a key role in ensuring that the best scientific and technical advice is available on a very broad variety of topics in the support of decision and policy making in different contexts. He/she will be recognised as a scientist of the first rank, able and acknowledged to exercise judgement impartially and objectively and to offer advice over the full range of the Department's business. The successful candidate will have considerable personal authority and will be an effective and persuasive communicator, able to contribute across Government.

This is a part-time (three to four days per week) appointment and carries a salary of up to £150,000 pro rata, dependent on track record and experience. It is a fixed term appointment for three years, renewable for up to a further two years at the agreement of both parties. This will be an appointment within the Senior Civil Service equivalent to that of a Director General. A secondment would be considered.

Defra is an equal opportunity employer.

Further information including details on how to apply can be obtained from Defra's Recruitment Consultants, **kmc international** at **www.kmcinternational.com** or by email from **defra@kmcinternational.com** or by telephone on **+44 (0)2920 353817.** Please quote reference number 1142/2.

The closing date for completed applications is Thursday, 29 March 2007.

defra
Department for Environment
Food and Rural Affairs

kmc
international

98228RM

DIRECTOR
Bigelow Laboratory for Ocean Sciences

Bigelow Laboratory for Ocean Sciences, West Boothbay Harbor, Maine seeks an imaginative, accomplished and entrepreneurial scientist with exceptional leadership and communication skills for the position of Director.

Bigelow Laboratory is an independent, non-profit organization renowned for its major contributions to oceanography through cutting-edge research on processes affecting the productivity of the oceans, coastal seas and estuaries. Founded by visionary scientists Drs. Charles and Clarice Yentsch in 1974, Bigelow Laboratory is distinguished by a spirit of scientific freedom, a tradition of open, interdisciplinary team-building, and mentorship. It is a unique community of scientists, staff, trustees and global neighbors with a common dedication to ocean sciences. Present interdisciplinary research strengths include phytoplankton ecology and evolution, optical oceanography, marine microbiology, and the ecology of higher trophic levels, with expansion into trophodynamics and ecological chemistry. Facilities include the Provasoli-Guillard National Center for Culture of Marine Phytoplankton, the MacIsaac Facility for Individual Particle Analysis and the Center for Transmission and Scanning Electron Microscopy. For more information, please visit http://www.bigelow.org

The Director will help Bigelow Laboratory realize a shared vision for future excellence, including planned expansion to a new facility. He/she will lead staff and operations, ensuring continuing recruitment, retention and development of top-quality staff. The Director will work effectively with the Board of Trustees, and will encourage creativity and innovation at all levels. She/he will help shape the agenda for future ocean sciences research at the state, national and international level; communicate the importance of oceanographic research to policymakers and the general public, and diversify Bigelow Laboratory's funding from public and private sources, building on its long-standing success in acquiring competitive external funding.

Salary will be competitive and commensurate with experience and qualifications; a complete benefits package is available. Nominations and applications, including curriculum vitae, evidence of leadership of multidisciplinary organizations, scientific achievement, and professional goals should be sent to the address below. All inquiries, nominations and applications will be treated confidentially. Review of applications will begin by April 1, 2007.

For more information, contact **Director Search Committee, Bigelow Laboratory, PO Box 475, W. Boothbay Harbor, ME 04575, USA** *or e-mail:* **Bigelowdirector@bigelow.org** *phone:* 207-633-9615

An Affirmative Action/Equal Opportunity Employer NW96738R

Change your environment. Find jobs where you'll make a difference

naturejobs